DOCUMENT RESUME

ED 056 889 SE 012 700

AUTHOR	Smith, T. C., Jr.
TITLE	The Utility of An Evaluative Model in Judging the Relationship Between Classroom Verbal Behavior and Student Achievement in Three Selected Physics Curricula, Final Report.
INSTITUTION	Houston Univ., Tex.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-9-G-044
PUB DATE	Aug 71
GRANT	OEG-7-9-530044-0115(010)
NOTE	126p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	*Comparative Analysis; *Course Evaluation; Critical Thinking; Doctoral Theses; Evaluation; Interaction Process Analysis; *Physics; *Secondary School Science

ABSTRACT

The purpose of the 1968-69 investigation was to determine the applicability of a curriculum evaluation model to investigate high school students' achievement in three physics courses (traditional physics, Physical Science Study Curriculum, and Harvard Project Physics). Three tests were used to measure student progress: The Dunning-Abeles Physics Achievement Test, Form E., the Wisconsin Inventory of Science Processes, and the Watson-Glaser Critical Thinking Appraisal, Form YM. The classroom verbal behavior of each class was recorded on audio tapes and quantified using Flanders' Verbal Interaction Analysis system. The results of the study indicate that the students exhibited greater change in physics content mastery than in understanding science processes or in critical thinking, although there was a significant increase in all three components. No specific curriculum was found [illegible]be more effective in enhancing student ability-growt[illegible] the three tested components (physics content, understanding science processes, and critical thinking). The total classroom verbal interaction patterns were found to differ among curricula and between high and low achieving classes, leading to the judgement that there is a relationship between classroom verbal behavior and a physics curriculum as well as a relationship between classroom verbal behavior and student achievement. The funds for this doctoral dissertation were provided by the U. S. Department of Health, Education, and Welfare. (Author/TS)

1-9-G-044

FINAL REPORT
Project No. 9-G-044
Grant No. OEG-7-9-530044-0115-(010)

THE UTILITY OF AN EVALUATIVE MODEL IN JUDGING THE RELATIONSHIP BETWEEN CLASSROOM VERBAL BEHAVIOR AND STUDENT ACHIEVEMENT IN THREE SELECTED PHYSICS CURRICULA

T. C. Smith, Jr.
University of Houston
Cullen Boulevard
Houston, Texas 77004

August 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 9-G-044
Grant No. OEG-7-9-530044-0115-(010)

THE UTILITY OF AN EVALUATIVE MODEL IN JUDGING THE RELATIONSHIP BETWEEN CLASSROOM VERBAL BEHAVIOR AND STUDENT ACHIEVEMENT IN THREE SELECTED PHYSICS CURRICULA

T. C. Smith, Jr.

University of Houston

Houston, Texas

August 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Acknowledgments

I wish to express appreciation to the many people and schools that have given their support and cooperation to this study. Without them, the study would not have been possible.

The personal support, encouragement, and academic guidance of Dr. Silas W. Schirner, chairman of the committee, I wish to acknowledge with deep gratitude.

The guidance of Dr. Jacob W. Blankenship, Dr. Robert E. McClintock, and Dr. Robert H. Walker, members of the research committee, is greatly appreciated.

For the energetic assistance of Dr. Robert E. McClintock, I will always be thankful.

Special thanks is expressed to Charlotte Oliver for her assistance in computer programming, audio tape analysis, and constant encouragement throughout the study.

Appreciation is extended to Mr. Philip O. Vogel and Mr. Alan J. Segal of the Bureau of Research at the University of Houston for their individual cooperation and assistance during the study.

The research presented or reported herein was performed pursuant to a grant from the United States Office of Education, Department of Health, Education, and Welfare. Grant number OEG-7-9-530044-0115-(010), Project number 9-G-044, was authorized under the Cooperative Research Act, P. L. 83-531, as amended by Title IV of P. L. 89-10, Sec. 2(a), Cooperative Research Program.

THE UTILITY OF AN EVALUATIVE MODEL IN JUDGING THE RELATIONSHIP BETWEEN CLASSROOM VERBAL BEHAVIOR AND STUDENT ACHIEVEMENT IN THREE SELECTED PHYSICS CURRICULA*

An Abstract of a Dissertation

Presented to

the Faculty of the College of Education

University of Houston

In Partia' F' .fill

of the Requirements for the Degree

Doctor of Education

by

T. C. Smith, Jr.

August 1971

*The research presented herein has bee reported to 'Jniversit; Microfilms, Ann Arbor, Michigan.

Smith, T. C., Jr. The utility of an evaluative model in judging the relationship between classroom verbal behavior and student achievement in three selected physics curricula. Unpublished doctoral dissertation, University of Houston, August, 1971.

Committee chairman: Dr. Silas W. Scharner

Abstract

The purpose of the present investigation was to determine the applicability of a curriculum evaluation model to investigate high school students' achievement in three physics courses (traditional physics, Physical Science Study Curriculum, and Harvard Project Physics). The model was based upon the premises that (a) evaluation should be viewed as a dua of description and judgment, (b) evaluation should contain a description of instruction and its relationship to student outcomes, (c) teaching is a process of interaction, (d) learning is a consequence due to the effect of classroom verbal behavior intersecting with a curriculum, (e) knowledge and learning consist of both content and process, and (f) these attributes can be described quantitatively.

The evaluation model was applied to a sample consisting of 954 students enrolled in 38 classes taught by 26 teachers. This sample was randomly selected from approximately 150 schools which offered one of the three selected courses during the 1968-1969 academic year, thus enabling inference to high school

physics curricula in the southwestern, midwestern, and western United States.

Formative experiences of the students--academic aptitude, age, curriculum, grade level, and sex--were assessed and necessary statistical controls applied. A battery of three tests were administered twice, once upon entry into the curriculum and once upon completion of the curriculum. The battery consisted of the following tests: the Dunning-Abeles Physics Achievement Test, Form E; the Wisconsin Inventory of Science Processes; and the Watson-Glaser Critical Thinking Appraisal, Form YM. The classroom verbal behavior of each class was recorded on audio tape four times throughout the 1968-1969 academic year, and subsequently quantified by application of Flanders' Verbal Interaction Analysis system.

A three-factor analysis of variance mixed design with repeated measures on two factors was used to analyze student achievement data for curricula, tests, and pre-post dependent variable dimensions. The classroom verbal interaction matrices were analyzed across curricula and as to the two extreme levels of student achievement. Nonparametric tests were employed as appropriate to determine differences in total classroom verbal interaction patterns, category differences, and differences among four selected aspects of classroom verbal behavior (indirect-direct ratio, teacher-talk ratio, student-talk ratio, and content ratio).

The secondary school physics students were found to be above average in academic ability and neither grade level nor sex of a student was a determining factor as to his level of physics achievement. The three curricula were judged to be equivalent and effective in implementing the learning of physics. The students exhibited greater change in physics content mastery than change in understanding science processes or change in critical thinking, although there was significant increase in all three components. No specific curriculum was found to be more effective in enhancing student ability growth in any of the three tested components (physics content, understanding science processes, and critical thinking).

The total classroom verbal interaction patterns were found to differ among curricula and the total classroom verbal interaction patterns also differed between high and low achieving classes. This led to the judgment that there is a relationship between classroom verbal behavior and a physics curriculum as well as a relationship between classroom verbal behavior and student achievement. Due to the problems involved in statistical isolation of the differences in patterns, a description of specific causal relationships was not possible.

The curriculum evaluation model was found to be valid and applicable in evaluating three physics curricula. However, such application was found to be hindered by the lack of precision in measuring instruments and non-robust statistical techniques for assessing key difference points within the classroom verbal interaction matrices.

Table of Contents

List of Tables

List of Figures

Introduction

During the past decade many new curricula have been developed for the sciences. These curricula were developed because of an increased awareness of needed reforms in the sequencing and structure of science instruction. In developing these new curricula, science educators raised many questions about the kinds, the methods, and the quality of courses taught in American high schools. The basic considerations in these new programs have centered upon selection of materials to teach and methods of presentation of the selected materials.

These new texts, materials, and instructional procedures have been introduced into science classrooms in an attempt to achieve more nearly the goals of science education. The introduction of these materials and procedures may be considered one of the critical factors in implementing change in science education.

Systematic evaluation is a means of gaining insight into the effectiveness of these curricula. A general evaluation model may be developed to aid in systematizing the evaluative process and to insure the consideration of key variables in complex teaching-learning environments. This study will apply such an evaluation model to a specific area of science education, physics. Evaluation of these curricula should reveal durable relationships that would serve as a guide for changing or modifying instructional programs in science.

Background of the Problem

The total public school enrollment in the United States has been increasing since 1890. During this same period of time, the percentage of high school students taking physics has been continuously declining while other areas of science such as biology, chemistry, earth science, and mathematics have not shown such declines in enrollment. For example, in 1890, 22.8% of the total high school enrollment was in physics; in 1933, only 8.9% was in physics; and in 1957, the percentage in physics was 4.5% (National Education Association, 1957). A further examination revealed that during the 1963-1964 school year only one in five twelfth-grade students was enrolled in any type of physics course (Watson, 1967). This trend of declining enrollment relative to total high school enrollment is a major concern of many educators. The reasons for such a trend have not yet been determined.

Brown (1939), one of the first educators to exhibit concern about the physics curricula of the early 1900's, criticized not only the unit concept, but also the unchanging nature of the units within the traditional physics (TP) course. These units were generally classified under the headings of mechanics, heat, electricity and magnetism, and sound and light (Beauchamp, 1933). Henry, in Science Education in American Schools (1947), stated: "The picture we get is a subject, physics, gone stale through adherence to a set and largely nonfunctional pattern of organization" (p. 209). The prevalent belief was that the content of physics was a verified and certain body of facts. As a result the teaching of physics was dominated by this belief. To facilitate the transmittal of facts, the classroom method of instruction was primarily lecture or lecture-demonstration followed by a directed type of laboratory exercise (Stollberg, 1960). Over the years, teaching in physics had become equated with "telling."

Minor revisions have been made in organization of the units of the TP course. What is referred to as a "traditional" physics course seeks to familiarize the student with virtually all aspects of physics. Emphasis on the role of the physics teacher has changed within the TP course from one of imparting knowledge by telling or lecturing to one of directing the learning process. These minor revisions have not resulted in marked changes in enrollment trends.

In 1956, a more radical revision of the physics curricula was proposed by the Physical Science Study Curriculum (PSSC) project, under the direction of Zacharias at Massachusetts Institute of Technology. PSSC attempts to present physics as a unified but continuing process by which man seeks to understand the nature of the physical world. Emphasis has been placed on relatively few concepts viewed as basic to the underlying structure of physics. Laboratory investigations were designed to encourage the students' spirit of inquiry. Diadactic classroom presentation was de-emphasized and inquiry teaching was encouraged. However, many educators (Watson; Rutherford, 1967; Holton, 1967) felt that PSSC was not meeting the needs of many students due to the depth of the conceptual development necessary for success in the course.

A second attempt to modify the physics curriculum has been through the efforts of Harvard Project Physics (HPP). HPP, under the direction of F. James Rutherford, Fletcher Watson, and Gerald Holton at Harvard University, developed a "second generation" physics course for national use. This course was developed for the purpose of providing a physics course with maximum flexibility of content, emphasis, and teaching methods. Emphasis was placed on independent study for maximum content mastery, process acquisition, attitudinal changes, scientific literacy, and career guidance (Watson, Rutherford, Holton). HPP was first introduced into physics classrooms during the 1968-1969 academic year. The effects of the curriculum are currently being investigated.

TP, PSSC, and HPP are being utilized concurrently in American high schools. PSSC and HPP have resulted in increased emphasis on student-centered activities in the classroom, resulting in a change in the role of the teachers from a "teller" to a resource person and in a change from highly structured classrooms to laboratory classrooms. The TP has also reflected similar changes in emphasis.

Although the general emphasis for the three curricula is the same, each one approaches implementation of the emphasis in a unique manner. The question of the relative effectiveness of the three curricula in reaching the common objective of providing a physics curriculum to meet the needs of today's young people is thus seen to be of utmost importance.

Statement of the Problem

The purpose of the investigation was to determine the applicability of a curriculum evaluation model in investigating high school students' achievement in three physics curricula (TP, PSSC, HPP). In applying the evaluation model, the common objectives from each curriculum are selected for measurement and decision making.

Four general questions are considered within the model:

1. Are there differences among composite student achievement change scores for the three physics courses?
2. Are there differences in student achievement among the three courses for each of the achievement measures?
3. Is there a relationship between classroom verbal behavior and student outcomes in each of the three courses?

4. Is there a combination of classroom verbal behavior categories and a particular physics course that produces maximum student achievement?

Description of an Evaluation Model

Rationale

The simple assertion has been made in educational theory that the function of the school is to change or develop behaviors. In studying the productions of these behaviors and the educational programs intended to produce them, both behaviors and programs must be viewed in terms of their constituent variables. The primary purpose of instructional theory is to make more precise the understanding of the relationship between educational environment (independent) variables and behavioral outcomes (dependent) variables, so that the independent variables can be manipulated to produce maximum values of the dependent variables (Ausubel & Robinson, 1969).

Evaluation's major role in developing a theory of instruction is to furnish information about relationships between the educational environment and behavioral outcomes. Cronbach (1963) has proposed that the main objective for instructional evaluation is to add description of instruction to the traditional description of pupil outcomes and to seek a description of relationships between them. Evaluation has been further viewed as a dual process that includes both description and judgment of a curriculum (Stake, 1969).

Although teaching and learning are distinctly different processes they are so closely related in the classroom that to describe and evaluate one while failing to investigate the other could seem an untenable approach to the development of an instructional theory.

The teaching-learning process is multidimensional and for a single study to examine all the dimensions would be an impossibility; therefore, a theoretical position must be assumed in viewing teaching and learning. Stake believes that: "In the matter of selection of variables for evaluation, the evaluator must make a subjective decision. Obviously, he must limit the elements to be studied. He cannot look at all of them" (p. 353).

In the present study a theoretical model for evaluation has been devised based upon the conviction that practice must have a theoretical basis and a valid theory must have supportive evidence in application. Therefore, from practice one can derive clarification and redefinition of theory (Van Dalen, 1959).

The evaluation model is based upon the premises that: (a) the teaching-learning process is an open, multifaceted system; (b) the major interactive components of the system are teacher behaviors, student behaviors, and curriculum; (c) each major component may be represented quantitatively by sampling specified behaviors or materials; and (d) evaluation is a dual process which consists of description and judgment of the system. The underlying bases of the model are the ideas for evaluation proposed by Stake and Biddle (1964).

The Model

The model consists of five major components (Figure 1). Each component is considered as interacting with the four other components in the dynamic environment of "reality." Theoretically, the five components may be sampled from and examined in a stable, independent state permitting inference as to the relationships of components when operating in "reality."

Figure 1 presents the five components (behavioral goals, formative experiences, transactions, outcomes, and judgments) separated logically in terms of the temporal sequence in which they occur. A simplified representation of the cause-and-effect sequence is indicated by the arrows connecting the components. Each of the various components will be discussed in order as they appear in the temporal sequence.

Components of the Model

Behavioral goals. Goals are statements which describe what the student should be able to do after completing an instruction sequence. The use of behavioral goals in evaluation of curricula has been advocated by many educators (Tyler, 1950; Bloom, Engelhart, Furst, Hill & Krathwohl, 1956; Mager, 1962; Glaser, 1965; Gagné, 1967; Popham, 1969).

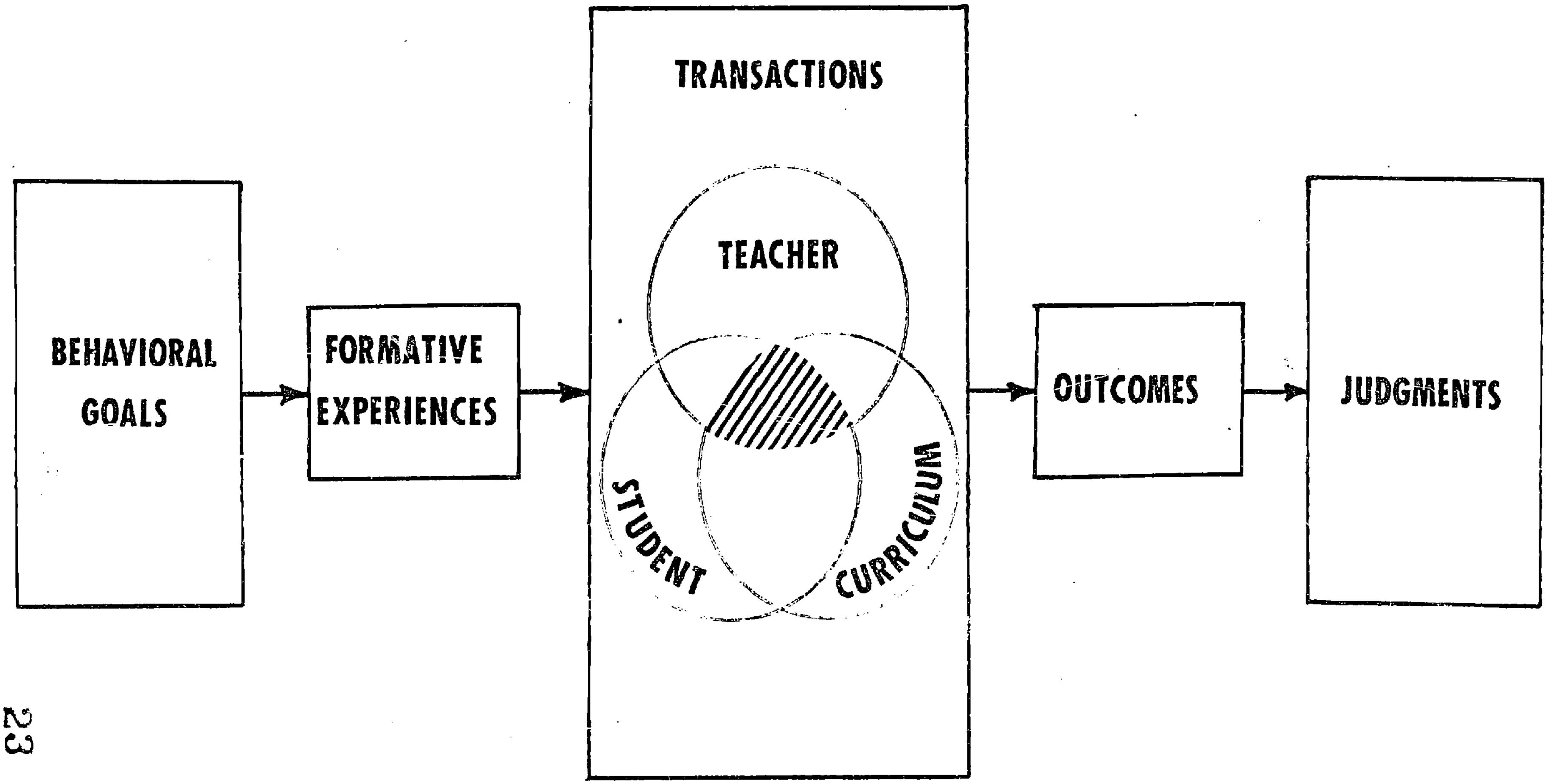

Figure 1

Five Component Evaluation Model

<u>Formative experiences</u>. A formative experience is any encounter or condition existing prior to teaching and learning which might relate to student outcomes. Ausubel & Robinson have pointed out the most important factors influencing learning are the quantity, clarity, and organization of the students' present knowledge prior to entry into a curriculum. This knowledge, available at any point in time, is referred to as his cognitive structure. Thurstone (1947) has shown that background characteristics of age, sex, intelligence quotient, and prior training or education are potential influences on students' achievement or outcomes. These characteristics cannot be expected to remain invariant from student to student or class to class. Thus some control must be utilized in evaluation to compensate for the differences within a sample under investigation.

<u>Transactions</u>. Transactions are the many encounters of teacher with students, students with students, teacher with subject matter, and student with subject matter. The interaction of these elements (teacher, student, and curriculum) form a triad which is dynamic in nature. This common concept of teaching as a process of interaction has been shared by Hughes (1959), Smith (1960), Bellack (1966), and Hyman (1967). Teaching is viewed in this dynamic triad of elements as the outcome of the interactive forces which produce student behavior change. Figure 2 shows the triad and the dynamic nature of the interactions.

Each element influences and is influenced by the relationship between the other two elements. The interaction between any two of the elements influences the mode of reaction of each of the two to the third one and, in turn, how all elements will react simultaneously.

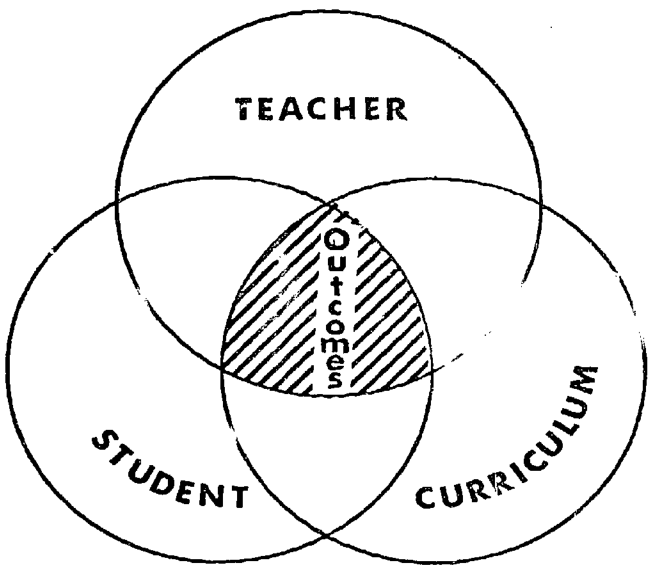

Figure 2

Teaching Triad

<u>Outcomes</u>. Outcomes are the measured consequences of an instructional system. These consequences may include a description of attitudes, cognitive development, role-perception, classroom climate, and social adjustment. Of the many attributes of the learning environment, the consequences selected for measuring must be isomorphic to the behavioral goals. The outcomes describe quantitatively the extent to which the educational goals are being attained.

Judgments. Judgments are decisions as to relative effectiveness of alternate approaches to attaining educational goals. Various aspects of the consequences may be compared and judgments formulated.

Of these five components, only the transactions stage may be viewed as dynamic. This dynamic stage connects the four relatively static components into an integrated sequence.

Application of the Model

Through the model, an attempt was made to measure the relative effectiveness of three physics courses (TP, PSSC, HPP) and the relationships of described classroom verbal behavior to student outcomes. Relative effectiveness among three physics curricula is defined as change scores between the pretest and posttest on selected criterion instruments.

Behavioral Goals

Two basic goals of science teaching were chosen for investigation because they are important cognitive skills expected in the student after completion of any curriculum or course (Stollberg; Victor & Lerner, 1967; National Science Teachers Association, 1970). These two basic goals are content of science and processes of science. Content of science surrounds acquisition of concepts, facts, theories, and principles. Processes include the many abilities, skills, and attitudes that make critical thinking and problem solving possible. Process is the means by which real learning of the content of science takes place in the science classroom (Victor & Lerner). From this point of view, processes are subsets of content learning and cannot be separated or dichotomized. To evaluate the effectiveness of a science curriculum, it is essential to measure change in both content and processes.

Formative Experiences

The major formative experiences were considered to be (a) length of time in the educational environment, (b) differential effects of cultural roles, (c) clarity and organization of cognitive structure, and (d) prior learning specific to the physics curricula. The criteria selected for these experiences were grade level, sex, academic aptitude, and content-processes pretests.

Transactions

The transaction component consists of three teaching-learning triads, each designated by a specific physics curriculum. Each triad consists of the variables within the classroom. Within each physics course the teacher has specific patterns of behavior which become interactive with students and their cognitive modes.

A teacher's method of instruction is a pattern or manner of treating students, classroom events, and subject matter. When the teacher and students interact, both use words to communicate ideas. The daily activities in the classroom are sustained almost entirely in talk between teacher and students (Aschner, 1961). Verbal communication is used in exchanging ideas about facts, concepts, principles, generalizations, and definitions. Smith (1956) contends that teaching cannot occur without the use of language. Similar positions have been taken by Bellack and Amidon & Flanders (1963).

Analysis of the physics classroom thereby consisted of systematic observation and quantification of the patterns of classroom verbal behavior. The most promising or logical approach to this quantification is that of Flanders (Belanger, 1964). Flanders' Verbal Interaction Analysis (FVIA) system (Amidon & Flanders) utilizes a ten category coding which relates to the social climate of the classroom. This recording system facilitates analysis of the antecedents of verbal acts and assists inference as to possible cause-effect relationships.

Outcomes

Measurements of outcomes are classified as to classroom climate, i.e., classroom verbal behavior and student cognitive behavior. Classroom climate variables are the ten categories of the FVIA. The student cognitive behavior variables consist of physics content attainment and level of functioning in science processes and critical thinking. Appropriate standardized tests were selected for measurement of each of these variables.

Judgments

The data generated by measuring the selected behaviors were analyzed according to three statistical procedures: the analysis of variance, Darwin's Chi-square, and a test of medians. Statistical decisions formed the bases for inference of relative effectiveness of each curriculum and the relationship of classroom verbal behavior to the inferred effectiveness. These inferences constituted the decisions produced by application of the evaluation model.

Review of Related Literature

Introduction

The purpose of this section is to summarize previous research which is related to this study. The reported studies are grouped into three categories: Science Curricula Studies, Physics Curricula Studies, and Classroom Verbal Behavior Studies. The latter is subsequently divided into a general section plus sections on student outcomes in science and student outcomes in physics courses.

Science Curricula Studies

Several studies in the sciences (chemistry, biology, and earth science) have focused on the new curricular materials to determine what effect the curriculum has on the students' achievement. A variety of criterion instruments have been used to measure the dependent variables in comparing the achievement of students in the new curricula with those in a more traditional course. The criterion instruments were usually administered on a pre- and posttest basis to measure change in some aspect of cognitive science knowledge.

A review of the literature indicates science educators have selected major components of cognitive science knowledge for evaluation in various ways. However, two components appear consistently throughout the literature: science content and science processes.

Content knowledge has been concerned with recall, recognition, application of concepts, facts, and principles specific to an area of science. Two types of standardized achievement tests are used to measure content knowledge. The first is to measure traditional science course materials (chemistry, biology, and earth science) and the second is to measure content specific to a curriculum such as Chemical Educational Material (CHEM) Study, Chemical Bond Approach (CBA), Biological Science Curriculum Study (BSCS), and Earth Science Curriculum Project (ESCP). Studies by Lisonbee (1964), Heath and Stickel (1963), Gennaro (1965), Kastrinos (1964), Lance (1964), Rainey (1965), Psychological Corporation (1965), Champlin and Hassard (1966), Schirner (1967), and Troxel (1968) have utilized one or more of these two types of tests to measure content knowledge.

Science processes are usually defined as understanding of scientific enterprise and critical thinking ability. Understanding of scientific enterprise includes scientific attitudes and assumptions and the methods and aims of science. The Test on Understanding Science (TOUS) by Cooley and Klopfer (1961) is one of the most frequently employed instruments in measuring this aspect (Schirner; Troxel; and Thomas, 1968).

The ability to think critically, a second major aspect of science processes, has been defined as the ability to recognize and define a problem, clarify a problem by making appropriate definitions, ability to distinguish between facts and assumptions, organizing pertinent information, formulating and testing hypotheses, and stating tentative conclusions. The _Watson-Glaser Critical Thinking Appraisal_ (WGCTA) (Watson & Glaser, 1964), has been used extensively as a measure of this aspect of science processes (Anderson, 1965; Kastrinos; George, 1965; Herron, 1966; Schirner; Troxel).

A variety of other instruments have been used to measure science processes (_Test on Science Knowledge_, _Cognitive Preference Test_, and Kastrinos' _Critical Thinking Test_). However, the selection of a particular instrument appears to be specific to the study (Kastrinos; Psychological Corporation; Champlin & Hassard; Schirner).

The design most frequently used in analyzing student data has been analysis of variance in which pretests were used as a control for prior science knowledge. Other entry skills or individual differences of subjects controlled were academic aptitude (IQ), sex, and grade level. Although the analysis of variance design prevails in science curriculum studies, one exception was noted. Herron used factor analysis to treat pre- and posttest data.

Conflicting trends appear in the results of the science curriculum evaluation studies reviewed. These conflicting trends appear to be a function of the use of two types of standardized achievement tests. When a traditional science achievement test is used in evaluating a new curriculum versus a traditional curriculum, the new curriculum results in as high or higher student achievement (Lisonbee; Gennaro; Kastrinos; Lance; Rainey; Marks, 1967; Troxel). When a specific curriculum achievement test (i.e., <u>ESCP Achievement Test</u>) is employed in evaluation, traditional curricula appear to be at a disadvantage (Heath & Stickel; Lisonbee; Gennaro; Psychological Corporation; Schirner).

One of the most consistent results revealed in the literature is that of change in the ability to think critically. Students in a new curriculum tend to score as high or higher on the WGCTA than students in a traditional course (Kastrinos; George, 1965; Marks; Schirner; Troxel; Thomas). When comparisons are made on achieving a greater understanding of the scientific enterprise the trend appears to be with the new curriculum (Marks; Troxel; Psychological Corporation; Thomas; Schirner).

Physics Curricula Studies

A search of the literature revealed several studies relating to physics achievement. The studies reveal the same trend for evaluation as the other sciences. The physics studies attempt to determine what effect the physics curriculum has upon the students' understanding of physics content, understanding of science, and critical thinking ability.

Several studies have evaluated the effect of a new physics curriculum versus a traditional course upon content achievement. These studies show that students in a traditional course achieve as high or higher on standardized achievement tests which were designed to test a broad knowledge of "classical" or traditional physics than students in a new curriculum. Furthermore, the literature shows that students enrolled in a new physics curriculum or course score higher on achievement tests designed specific to that curriculum. Sawyer (1963) investigated this trend while attempting to measure the effectiveness of the Physical Science Study Curriculum (PSSC) program relative to traditional physics. A test was developed which included content items from both PSSC and non-PSSC materials. He concluded that PSSC program does not fulfill non-PSSC objectives and vice versa, and that the tendency for students taking such a combination examination would favor higher scores for non-PSSC students.

Heath (1964) reported a study which found that traditional high school physics classes achieved higher on a conventional test than PSSC classes. However, the PSSC classes performed in a superior manner on the PSSC Final Examination than did the traditional classes. A similar evaluation study was reported by Berry (1966). The Cooperative Physics Test and the PSSC Final Examination were used to test two groups. No significant difference in achievement was found between the groups, however students in the PSSC physics course learned more of the essential subject matter content of PSSC physics than did students enrolled in the traditional physics (TP) course.

Welch and Rothman (1968) conducted an evaluation of Harvard Project Physics (HPP) in which changes in physics achievement were measured for two groups. One group consisted of students who voluntarily enrolled for HPP and the other group consisted of students who were requested to register for HPP. There were no significant difference between the two groups as measured by the Physics Achievement Test for HPP.

Niman (1970) developed and tested a set of materials that explored sophisticated physical ideas through the use of mathematical models of physics for teaching. To measure physics content, a final examination was prepared and administered to an experimental group and a control group. The experimental group was taught using the developed mathematical models. The study revealed that students in the experimental group performed significantly better on the physics test than the control group.

The literature reveals that critical thinking is a major aspect of science processes to be evaluated in physics curricular studies and that the WGCTA is the instrument used to measure this aspect. Three studies have shown that students enrolled in PSSC physics develop the ability to think critically to a greater extent than students of a TP course. Brakken (1965) factor analyzed tests and compared two groups in critical thinking. He found that both TP and PSSC increased critical thinking, but PSSC accomplished it to a greater degree. Day (1964) and Rutledge (1965) also investigated change in critical thinking of PSSC and TP students. Both reported that students enrolled in PSSC physics developed into significantly better critical thinkers than those in a TP course.

The TOUS and Science Process Inventory (SPI) (Welch & Pella, 1968), are the most frequently used instruments for assessing change in understanding of science as a result of physics curricula. Two studies (Crumb, 1965; Trent, 1965) have investigated PSSC and non-PSSC students' understanding of science as measured by the TOUS. Crumb compared students taught by PSSC physics teachers, both with and without formal PSSC training, and teachers of non-PSSC physics, both with and without formal PSSC training. Results indicated that PSSC physics students score significantly higher than students in the non-PSSC course. Trent, using a different sample of PSSC and non-PSSC students, found no significant difference between PSSC and non-PSSC students' achievement as measured by TOUS. Both studies reported that adjustments were made for scholastic aptitude and prior science experience. Since the same student variables were controlled, it appears the difference could be attributable to uncontrollable teacher factors.

Welch and Rothman, mentioned previously, used both the TOUS and SPI to measure change in understanding of science.

The study compared two groups within a new physics curriculum, HPP. No significant differences were found between groups. Niman, also mentioned earlier, employed the SPI to measure understanding of science. There was no difference in achievement between experimental and control group as measured by the SPI.

A search of the literature on physics curriculum studies indicates that various comparisons have been conducted to determine the effect of traditional physics versus PSSC, or TP versus an experimental course. The selection of components for evaluation are the same as those revealed in other science curriculum studies: content and processes. Several of the studies reviewed compared the effects of two curricula, TP and PSSC, but none reported an analysis of TP, PSSC, and HPP with respect to physics content and science processes.

Classroom Verbal Behavior Studies

Researchers in education have recently begun to study the interactive verbal patterns of behavior of pupils and teachers in the classroom. Classroom verbal behavior constitutes the verbal discourse or talk which takes place between teachers and pupils and between pupils and pupils. Several observational systems designed to analyze pupil-teacher interaction in the classroom have been developed. The earliest observational studies of classroom interaction began with the work of Thomas &

others (1929); followed by those of Anderson (1939); Lewin, Lippitt & White (1939); and Withall (1940).

Flanders (1960), who was influenced by the above mentioned research, developed concepts of directness, indirectness, and flexibility to describe classroom verbal behavior. These concepts had their inception from social psychological theory. Flanders' system is designed to test the effect of social-emotional climate on student attitudes and learning. The system consists of ten categories for the analysis of teacher pupil verbal interaction. Flanders used the observational system to study teacher influence styles, pupil attitudes, and resulting achievement in seventh-grade social studies and eighth-grade mathematics. This study suggests that teachers who provided flexible influence styles from direct to the indirect depending on the situation were better able to create climates which enhanced student achievement. The students of those teachers who were less flexible in their teaching achieved at a lower level.

Classroom Verbal Behavior and Student Outcomes

Since the early study by Flanders, several studies have investigated the relationship of patterns of teacher influence to pupil achievement in the schools.

Furst (1965), using interaction analysis, found that above average student achievement was positively related to indirect teacher influence, a moderate pace of teacher-pupil interaction, and an indirect teacher response to student talk. The study further revealed that the amount of student talk was positively related to student achievement.

Perkins (1965), in a study concerned with under-achievement with high-ability fifth graders, found that teacher lecturing and criticizing was related to loss in reading comprehension as measured by the California Achievement Test. Soar (1966) in a similar study, found that indirect teaching produced greater growth in reading comprehension in elementary school pupils than direct teaching. Powell (1968) found the same trend when investigating the students' arithmetic achievement.

Classroom Verbal Behavior and Student Outcomes in Science Curricula

Science educators have recently recognized that the teacher's behaviors determine in part the favorable teaching-learning environment. The rapid increase in the number of science teacher behavior studies and interaction studies in the last five years is evidence of this increasing awareness. A survey of the teacher behavior and interaction research reveals that many researchers are investigating the relationship between descriptions or patterns of teacher behavior to criterion measures such as pupil attitudes and pupil achievement.

Parakh (1968) used a modified Flanders Verbal Interaction Analysis (FVIA) category system to describe verbal patterns of biology teachers. He found a relatively low percentage of student verbal participation, especially student initiated contribution, compared with a high percentage of direct verbal teaching procedures used by most of the teachers in the study. The investigator did not attempt to make any evaluation statements about the teaching behavior patterns observed.

In another biology research project Gold (1966) investigated the verbal behavior of selected biology teachers. The teachers effectiveness was determined by scores on three instruments. The instruments were Teacher Rating Scale, Student Opinion Questionnaire, and Teaching Situation Reaction Test. The verbal interaction of the classroom was classified by a sixteen-category system developed by Hough (1967). Results indicated that the overall patterns of classroom behavior for two selected groups were not the same.

Schirner used FVIA to ascertain the directness or indirectness (ID ratio) of a teacher's pupil-teacher interaction in various teaching activities in earth science classes (ESCP and non-ESCP). To establish effectiveness, he used measured student outcomes on a set of criterion instruments. These instruments included TOUS, WGCTA, Test of Science Knowledge: Parts I and II, ESCP-Final, and Earth Science Final. He also measured the teacher's expressed philosophical orientation (TNT ratio). A Credo Preference Check List was constructed to determine the TNT ratio. The findings indicated the compatibility of the factors represented by ID ratios, TNT ratio, and the type of course led to significantly greater achievement in most student outcomes. A student who has a teacher that is direct and has traditional beliefs has an advantage if in a non-ESCP course and is at a disadvantage if in an ESCP course. If a teacher is indirect and has nontraditional beliefs, a student has an advantage if in an ESCP course and is at a disadvantage if in a non-ESCP course.

Evans (1968) investigated the relationship between teacher verbal behavior and pupil achievement over a three-year period with one teacher, and over another year period with the same pupils under a different teacher. The teacher behaviors were quantified through the FVIA system. The sample consisted of elementary children. He concluded that pupils under indirect teachers scored higher on Science Research Associates Achievement (SRA) tests than did pupils after three years of direct teaching. Examination of fourth-year scores indicates that some pupils, after three years of direct teaching, gained more in the fourth year than did those pupils with three years of indirect teaching. The author concluded that the measured differences in classroom climate were not related to pupil achievement and if the study had been restricted to the first three years of schooling then indirect teaching facilitates pupil achievement.

Citron & Barnes (1970) designed a study to gather statistical evidence to prove or disprove whether or not certain patterns of teaching were better than others for slow learners in biology at secondary school level. Achievement was measured in three areas: concept formation, problem solving, and total performance. The classroom observation instrument used to measure and verify prescribed patterns of teaching was the FVIA. The curriculum applied to all of the classes in the study was "Patterns and Processes" published by BSCS. The tests used to measure achievement were unit tests used with "Patterns and Processes." The conclusions reached from the study include: (a) a high indirect to direct teaching ratio (high I/D) early in the course for slow learners followed by a lower I/D ratio later in the course increases achievement in problem solving and in total performance, provided problem solving plays a major role in total performance; and (b) a constant intermediate I/D ratio throughout a course of biology for slow learners leads to a higher achievement in concept formation than does a change in I/D ratio in either direction from the start to the end of the course.

Evans & Balzer (1970) developed a category system based upon actual descriptions of teacher behaviors and then used the coding system to obtain an objective description of the classroom behaviors of a sample of biology teachers. The instrument, Biology Teacher Behavior Inventory (BTBI), was developed from video tape recordings of eleven biological science teachers. An analysis of the data revealed that when the behaviors of all the teachers taken together, over 94% of them pertained either to management or content development of the biology class.

Classroom Verbal Behavior and Student Outcomes in Physics Curricula

A search of the literature revealed two relevant studies involving the classroom behavior of the physics teacher and their students. The primary concern of these two studies was to relate behavior patterns to teaching effectiveness.

In the first study, Snider (1966) investigated a sample of 17 physics teachers in the New York State Regents Physics Course and their students. Flanders' method was used to describe physics teaching in terms of pupil-teacher verbal interaction on the directness of teacher influence dimension and to seek relationships between teacher effectiveness and both directness and flexibility of teacher influence. The teaching activities were segmented into Planned Demonstration, Lecture, Laboratory and Recitation-Discussion. Three selected aspects of student achievement were measured with paper-and-pencil tests. The results showed that matrix measures were, in effect, measures of teacher style and that style varied from one major activity to another. He found that the employment of social skills connected with aspects of positive motivation is an uncommon phenomenon in the high school physics classroom, physics teachers rarely build on student ideas, and there appeared to be relatively little learning through discovery in physics classrooms. Furthermore, no single directness or flexibility measure appeared as a factor of teacher effectiveness for all aspects of effectiveness considered. This dependent variable of effectiveness was inferred from the achievement measures.

In a second research project, Pankratz (1966) investigated a population of 30 physics teachers from 30 separate high schools in Columbus and Dayton, Ohio. A Teacher Rating Scale, Student-Opinion Questionnaire, and a Teaching Situation Reaction Test were instruments used to measure teacher effectiveness. The five highest and five lowest ranking teachers according to the three evaluative instruments comprised a high and low sample which were observed. Using the Observational System for the Analysis of Classroom Instruction developed by Hough the researcher classified and recorded verbal classroom behavior. The following conclusions were drawn from the data analyzed: (a) teachers who rated high on the evaluative instruments used more praise and reward, more cognitive and skill clarification and acceptance than teachers rated low; (b) indirect influence as compared with direct influence was employed by the high group significantly more often and in a more sustained manner; (c) the sustained use of student's ideas and the length of teacher's answers to students' questions was greater for the high rated group, (d) each of the two groups used approximately one-half of their time in lecturing; and (e) the high rated group used more indirect influence patterns, whereas the low sample emphasized direct influence patterns to a greater degree.

Summary

A survey of the literature relating to science (chemistry, biology, and earth science) curricula and specifically physics curricula leads to several general conclusions. Science educators have selected two components of cognitive science knowledge for evaluation. The two components of science content and science processes appear consistently throughout the literature. These two components selected for measurement are the same as the common objectives of science instruction identified by the model. In measuring these objectives, conflicting trends appear in student outcomes. The conflicts appear to be a function of the use of two types of standardized achievement tests used to measure science content and science processes. When a traditional achievement test is used in evaluating a new curriculum versus traditional curriculum, the new curriculum results in as high or higher student achievement. When a specific curriculum achievement test is used in evaluation, the traditional curriculum appears to be at a disadvantage. Another factor which may contribute to the conflicting trends is lack of uniform control of entry skills. These findings tend to support the necessity of the formative experiences component of the evaluation model.

Other studies stressed the importance of the teachers' verbal behavior in the teaching-learning environment. Several studies in science have investigated the relationship between patterns of teacher verbal behavior in various curricula and pupil achievement.

Results would indicate that science teachers, as a group, tend to be more direct than indirect in their classes and students' verbal participation is relatively low when compared to the teacher's. Trends further indicate teachers spend the majority of class time in content development. The effect of such behavior has been revealed in two physics studies. The two studies agreed that physics teachers rarely build on student ideas, that there is relatively little learning through discovery, and that nearly one-half of the class time is spent in lecturing. Conflicting results were reported with respect to the relationship of student achievement with teacher directness or indirectness. This conflict may have resulted from the restricted sample of the studies.

The present study investigated this conflicting and relatively unexplored area of verbal behavior of teachers within three physics curricula (TP, PSSC, and HPP) and the relationship to student outcomes. The transactions component within the evaluation model represents this relationship through the use of the student-teacher-curriculum triad. The following section will identify the method and procedures for selecting and measuring key variables within the various components of the evaluation model.

Method: Application of the Model

The purpose of this section is to present a description of the method used in applying the evaluation model to the physics curriculum. This application includes the following components: selection and description of sample, selection and description of measuring instruments, selection and description of an observational instrument, and the procedures of collection and analyses of data.

Selection and Description of the Sample

The subjects used in this study were 26 teachers and 954 pupils selected from approximately 150 schools which taught a traditional physics (TP) course, Physical Science Study Curriculum (PSSC), or Harvard Project Physics (HPP) course during the 1968-1969 school year. In the summer of 1968 a letter [Appendix A], including a questionnaire, was sent to approximately 600 Texas schools and National Science Summer Institutes across the United States. From the responses, only those schools that taught the TP, PSSC, or HPP curriculum were considered for sampling. The random sample consisted of 24 schools throughout 7 states with 38 classes containing 954 physics students in grades 9 through 12. Table 1 shows the number of schools in each physics curriculum, the number of students of each sex and the total number of teachers and students in the study. [A listing of the schools and their locations is included in Appendix B.]

TABLE 1

Numerical Description of the Sample

Physics Curriculum	Students			Teachers
	Males	Females	Total	
TP	195	52	247	8
PSSC	289	94	383	10
HPP	241	83	324	8
Total	725	229	954	26

Measures of academic aptitude were secured from each student's cumulative record on file in his school in order to eliminate the necessity of administering an additional test during the study. Since only the most recent measure on each student was used, a total of 13 different test scores were used. Those tests used to measure academic aptitude (IQ) were the following: (a) Differential Aptitude Test, (b) Preliminary Scholastic Aptitude Test, (c) American College Test, (d) College Entrance Examination Board, (e) Test of Academic Progress, (f) Otis, Gamma, (g) Iowa Test of Educational Development, (h) Henmon-Nelson Test of Mental Abilities, (i) National Merit Scholarship Examination, (j) Ohio State Psychological Examination, (k) California Test of Mental Maturity, (l) Science Research

Associates Test Battery, and (m) Lorge-Thorndike Intelligence Test. Verbal and numerical scores were secured from those tests that did not render a composite score; such scores were then converted to composite standard form to provide a basis for comparison. The raw scores from the 13 separate tests were converted to stanine form by using a normal pe... ile chart. To be directly comparable, each of the 13 tests should have their stanine conversions based on the same pupil population. However, each test had been standardized over representative populations of pupils and therefore should be comparable for a large population. Stanines are standard scores having a mean of five and a standard deviation of two. While stanine scores lack the precision of some other types of standard scores, it was judged that the compilation of different test results for different students would make this less precise form of score the most appropriate for this study.

The sample, while typical of the population of high school physics students, is not typical of the high school population as a whole. The sample in this study had a mean stanine of 6.71 with a standard deviation of 1.58. The academic aptitude of the sample is somewhat higher and less variable than that of the high school population at large; this finding was not unexpected.

Though all pupils were tested six times during the year (three pre- and three posttests) relatively little information was missing on any one student. Complete data were available for about 96% of the total sample. The mean of each variable for each class was computed and inserted in the place of a missing value. It was judged that approximately 4% missing data, due to absences, withdrawals, and other causes, was relatively low for a study of this type.

Selection of Tests

Three criterion instruments were selected in an attempt to examine different types of learning: physics content achievement, understanding science processes, and critical thinking. The Dunning-Abeles Physics Achievement Test, Form E (DAPT) (Dunning & Abeles, 1967) was used to measure physics content and concepts. It attempts to measure physics course content and concepts representative of any curriculum, new or traditional. The test contains 50 multiple choice questions; its reliability coefficient of 0.87 was determined by the split-half method as reported in its manual. Form E of the test has a mean of 25.5 and a standard deviation of 8.6. The test purports to measure five content and concept areas of physics; these areas are mechanics, electricity and magnetism, atomic and nuclear physics, wave motion and light, and kinetic-molecular theory. There are questions concerned with knowledge, understanding, and application in each of the five areas.

The Wisconsin Inventory of Science Processes (WISP) was used to measure knowledge of the scientific enterprise. The test items are concerned with the assumptions, activities, objectives, and products of science. The test consists of 93 multiple choice questions and has a mean of 54.2 for a sample of twelfth-grade high school students. The instrument has a reliability of 0.82 using Kuder-Richardson For ula 20, as reported by the Scientific Literary Research Center at the University of Wisconsin.

The Watson-Glaser Critical Thinking Appraisal (WGCTA), (Watson & Glaser, 1964) consists of a series of test exercises which require the application of some of the important abilities involved in critical thinking. The test contains 100 multiple choice questions. It contains five subtests designed to measure aspects of critical thinking. The WGCTA, Form YM, has an odd-even split-half reliability of 0.87. This coefficient was corrected by the Spearman-Brown formula. The five subtests of the WGCTA are (a) Inference (20 items), (b) Recognition of Assumptions (16 items), (c) Deductions (25 items), (d) Interpretation (24 items), and (e) Evaluation of Arguments (15 items).

Four other variables describing formative experiences were also included in the analysis: academic aptitud , sex of the student, curriculum, and grade level of the student.

Observational Instrument

The Flanders Verbal Interaction Analysis (FVIA) system utilizes a ten category scheme for classifying interaction data.

The ten categories can be grouped into three broad divisions: (a) teacher-talk, (b) student-talk, and (c) silence or confusion. Teacher-talk is divided into two types of influence, indirect and direct. The indirect influence consists of four categories:

1. Accepting feeling,
2. Praising or encouraging,
3. Accepting or using ideas of students, and
4. Asking questions.

The direct influence consists of three categories:

5. Lecturing,
6. Giving directions, and
7. Criticizing or justifying authority.

The student-talk is divided into two categories:

8. Student response, and
9. Student initiation.

Category 10 is classified as silence or confusion and defined as those pauses of silence and periods of confusion in which the verbal behaviors cannot be understood by an observer or coder. [Descriptions of the scoring categories and a sample Interaction Matrix Form are shown in Appendix C.]

Flanders developed a matrix technique which allows for the preservation of the sequential nature of the data. In the matrix, the rows and columns are numbered according to the numbers of the categories of the Flanders system described above.

Each numeral recorded on the coding sheet (an event or category is recorded every three seconds) is paired with the numeral immediately following it and the number pair tabulated into the matrix, the first number of the pair indicated the row and the second number the column into which the pair of numerals will go.

Procedures

Administration of Tests

Pretests of the DAPT, WGCTA, and WISP were administered to the subjects in September, 1968. Posttests using the same instruments were administered in May, 1969.

The teachers within the various physics curricula administered the criterion battery of tests during regular class periods in which a three-hour period of time was utilized. Each teacher received identical written instructions pertaining to the testing procedures to follow during test administration. The tests were sequenced in the same manner for all classes during the pre- and posttesting. All subjects were tested using the three measures for determining course effectiveness.

Measuring Classroom Verbal Behavior

Classroom verbal behavior which occurred in the physics classrooms of the subjects was recorded on audio tape to permit analysis of the ongoing verbal behavior. Regular classes were recorded four times during the 1968-1969 academic year for subsequent coding using the FVIA. The tapes were analyzed using the ten categories previously listed; the pupil-teacher verbal behavior was sampled and recorded at three-second intervals. This procedure produced sequences of numbers that represented the original sequences of verbal events or interaction of categories that took place in the classroom. After each tape had been coded, the numbers were tabulated in a 10 x 10 square matrix. Individual classroom verbal interaction matrices were combined to form composite matrix of multiple classes; matrices of the classrooms were compared in two ways: first, composite verbal interaction matrices were formed--one each for the TP, the PSSC, and the HPP classes; and second, composite verbal interaction matrices were formed from the highest and lowest scoring classes based upon composite achievement. The highest and lowest classes were determined by selecting the highest and lowest 27% of all classes in the sample, thus yielding maximum discrimination (Kelly, 1939). These two sets of composite matrices were then compared on an intra-set basis to assess any differences that might exist in the overall interaction pattern for the ten categories of behavior.

A comparison was made among the three curricula using only the highest scoring classes in an effort to assess differential patterns for classrooms showing the highest composite achievement. Comparisons were also made of individual classroom matrices for the three curricula and for the total high versus low scoring classes to determine if there were any differences with respect to the following.

1. The percentage frequencies in each of the ten categories.
2. The indirect-direct (I/D) ratio.
3. The teacher-talk ratio.
4. The student-talk ratio.
5. The content ratio.

Computational formulas used to compute individual values for the comparisons of ratios are given in Figure 3.

Ratio	Computational Formula
Indirect-direct ratio =	$\frac{\text{Sum of tallies in columns 1 thru 4}}{\text{Sum of tallies in columns 1 thru 7}}$
Teacher-talk ratio =	$\frac{\text{Sum of tallies in columns 1 thru 7}}{\text{Total number of tallies in matrix}}$
Student-talk ratio =	$\frac{\text{Sum of tallies in columns 8 thru 9}}{\text{Total number of tallies in matrix}}$
Content ratio =	$\frac{\text{Sum of tallies in columns 4 thru 5}}{\text{Total number of tallies in matrix}}$

Figure 3

Formation of Classroom Verbal Behavior Ratios

Establishing Coder Reliability

Previous studies by Flanders and others have shown that data gathered by means of interaction analysis are no more valid than the reliability of the coders. Prior to analyzing the tapes recorded in the classrooms, the two coders used Flanders' training tapes containing classroom sessions which provided variability of categories. They trained approximately ten hours before coding the tapes used in this study.

A method for establishing inter-coder reliability was developed by Scott (1955). Scott calls his reliability coefficient "pi" (π) and it is determined as

$$\pi = \frac{P_o - P_e}{1 - P_e}, \text{ where}$$

P_o is the proportion of agreements between coders and P_e is the proportion of agreement expected by chance, which is found by squaring the proportion of tallies in each category and summing those over all categories, such that

$$P_e = \sum_{i=1}^{k} P_i^2, \text{ where}$$

P_i is the proportion of tallies falling into each of the k categories. π can be expressed in words as the amount that two coders exceed chance agreement divided by the amount that perfect agreement exceeded chance. A Scott π of 0.85 or higher has been designated as a reasonable level of performance.

In this study a coefficient of 0.98 was established between the coders; data on which π was based are presented in Table 2.

TABLE 2

Calculation of Coder Reliability Using the Scott Method

Category (1)	Observer A (2)	Observer B (3)	% A (4)	% B (5)	% Diff. (6)	$(Av. \%)^2$ (7)
1	0	0	0.0	0.0	0.0	0.000
2	29	29	8.4	8.4	0.0	0.706
3	5	5	1.4	1.4	0.0	0.020
4	57	59	16.5	17.1	0.6	2.822
5	140	138	40.6	40.0	0.6	16.241
6	1	1	0.3	0.3	0.0	.001
7	0	0	0.0	0.0	0.0	0.000
8	43	43	12.5	12.5	0.0	1.563
9	41	41	11.9	11.9	0.0	1.416
10	29	29	8.4	8.4	0.0	0.706
Total	345	345	100.0	100.0	1.2	23.475

$$\text{Scott's Coefficient } \pi = \frac{P_o - P_e}{100 - P_e} = \frac{(100-1.2)-23.475}{100-23.47} = 0.98$$

Statistical Techniques

A 3 x 3 x 2 mixed factorial design (curricula, tests, pre-post) with repeated measures on two factors was used to analyze the pre- and posttest data. To satisfy the basic assumptions of the analysis of variance (ANOVA) design, Hartley's $F_{max.}$-test (Winer, 1962) was used to test for homogeneity of variance. All test scores for each criterion instrument were converted to standard scores to obtain equivalent scaling for the analysis of data. An XDS Sigma-7 computer, using a 3 x 3 x 2 mixed factorial design written specifically for the study, was employed in the analysis. The Scheffé a posteri tests (Winer) were used to determine between which means there was a significant difference. Correlations between the variables of grade level, academic aptitude, sex, and test scores (pre-post) were computed to measure their degrees of interrelationship.

The Darwin Chi-square test (1959) was used to compare the total FVIA patterns for the three physics curricula. All of Darwin's analysis is based on the assumption that interaction sequences are one-dependent, or Markov chains, which is a better approximation than the zero-dependent assumption of the Chi-square. Darwin contends that communication events are, in fact, more than one-dependent, but the additional dependence of three or more events is small by comparison to the dependence between two events. The Kruskel-Wallis H-test and Mann-Whitney U-test (Siegel, 1956) were used to make comparisons between variables of the FVIA matrices. All tests of significance were made at the .05 level of significance.

In summary, the statistical techniques employed in the present study included the following:

1. Pearson's product-moment correlation to measure the relationship between two variables.

2. ANOVA, three-factor mixed design with repeated measures on two factors (3 x 3 x 2 design).

3. Hartley's $\underline{F}_{max.}$-test for homogeneity for variances.

4. Scheffé method for determining significan differences between means.

5. Darwin Chi-square test for determining significant differences among total FVIA patterns of the three physics curricula.

6. Kruskel-Wallis $\underline{H}$-test for determining differences between variables of the FVIA matrices in comparing three groups.

7. Mann-Whitney $\underline{U}$-test for determining differences between variables of the FVIA matrices when comparing two groups.

Hypotheses

The following specific null hypotheses were tested:

I. There are no significant differences among composite achievement scores for the three physics curricula.

II. There are no significant differences in achievement among the three curricula for each of the achievement measures, i.e., no significant statistical interaction exists between curriculum and tests.

III. There are no significant differences among composite verbal interaction matrices for the three curricula with respect to the total interaction pattern for all ten categories.

IV. There are no significant differences among the classroom verbal interaction matrices for the three curricula with regard to:

1. The percentage frequencies in each of the ten categories.
2. The Indirect-Direct ratio.
3. The teacher-talk ratio.
4. The student-talk ratio.
5. The content ratio.

V. There is no significant difference between the high and low scoring classroom composite verbal interaction matrices with regard to the total interaction pattern for all ten categories.

VI. There is no significant difference between the classroom verbal interactions matrices for the high and low scoring classes with regard to:

1. The percentage of frequencies in each of the ten categories.
2. The Indirect-Direct ratio.
3. The teacher-talk ratio.

4. The student-talk ratio.

5. The content ratio.

VII. There are no significant differences among the classroom verbal interaction matrices for the high scoring classes of the three curricula with regard to:

1. The percentage frequencies in each of the ten categories.

2. The Indirect-Direct ratio.

3. The teacher-talk ratio.

4. The student-talk ratio.

5. The content ratio.

Summary

The evaluation model was applied to a sample consisting of 954 students enrolled in 38 classes taught by 26 teachers. This sample was randomly selected from approximately 150 schools which taught TP, PSSC, or HPP during the 1968-1969 school year. The sample may therefore be considered representative of the three curricula.

The behavioral goals component within the model identified two major areas of knowledge to measure: physics content and science processes. These goals were measured by a battery of three tests (physics content, science process, and critical thinking) selected and administered twice, before and after an interval of time, to three groups. Formative experiences, the second component, were included in the student information and subsequently in the analysis of data. Academic aptitude, sex of the student, curriculum, grade level, and pretest scores were the controlled entry skills for each student.

Within the transactions component, containing the teaching triad, classroom verbal behaviors of students and teachers within the three curriculum were quantified using the FVIA. The classroom verbal behavior patterns were then analyzed and related to composite (content, processes, and critical thinking) student outcomes, the fourth component of the model.

Based upon the analysis of data, judgments were formulated concerning the effects of classroom verbal behavior upon student outcomes within the three curricula. The following section presents the data analysis, results, and interpretations.

Data Analysis, Results, and Interpretation

The purpose of the previous section was to present a description of the method used to apply the model; following that, the present section presents the analyses of data, results, and interpretations. The section is divided into three major divisions: analysis of test data, analysis of classroom verbal behavior, and summary. Within each analysis section, the relevant hypotheses are discussed.

Analysis of Test Data

The principal method of test data treatment was the analysis of variance. In order to form a composite score for the three tests, it was necessary to convert the raw data from pre- and posttesting into standard scores. [The class raw score means are presented in Appendix G.] A simple z-transformation was performed to achieve equivalent scaling across the three instruments since each had a different mean and variance. The standard scores resulting from the transformation had a mean of 50 and a standard deviation of 10.

To satisfy the basic assumptions of the design, Hartley's F_{max}-test (Winer, 1962) was used to test for homogeneity of variance. The *F*-ratio of test variances between the curricula was less than the critical value of the .05 level; therefore, the data were assumed to meet the requirement of equivariance. Since the data showed homogeneity of variance, a test for normality of distribution was not conducted. Results reported by Box (1954) and Norton (Lindquist, 1953) have indicated that for unequal sample sizes, the data should be tested for equivariance and, if this proves tenable, the analysis of variance may be continued without further tests of assumptions.

The data were then treated to determine if formative experiences or entry skills of the students were significantly different for the several levels of the dependent measures. The variables of grade level, academic aptitude, and sex of the student were correlated with test scores, using the Pearson product moment correlation technique. [These correlation coefficients are presented in Appendix F.] Results indicated that test change scores could not be attributed to formative experiences. Since the formative experiences were not significantly different, the data were analyzed using the analysis of variance, the results being summarized in Table 3. It may be noted that four of the mean square ratios were found to be significantly higher than chance expectancy at the .05 level, resulting in significant *F*-values. Since the interaction of the tests and the pre-post dimension were found to be significant, the lower order main effect for the pre-post dimension is included in the interactive comparison and was not subject to additional analysis.

TABLE 3

Summary of Analysis of Variance for Student Outcomes on Evaluation Instruments

Source of Variation	SS	df	ms	F
Total	635,904	5723		
Between Subjects	(310,712)	(953)		
Curricula (C)	9,695	2	4847	15.315*
$Error_b$	301,017	951	317	
Within Subjects	(325,192)	(4770)		
Tests (T)	0+	2	0	
Pre-Post (PP)	28,818	1	28,818	536.361*
Interactions				
C X T	361	4	90	1.193
C X PP	451	2	226	4.197*
T X PP	25,165	2	12,583	318.561*
C X T X PP	236	4	59	1.494
$Error_{w1}$	143,940	1902	76	
$Error_{w2}$	51,096	951	54	
$Error_{w3}$	75,125	1902	39	

*$p < .05$

+Between-test variability removed during rescaling process.

Relative to Hypothesis I, significant differences among composite achievement change scores for the three physics courses were found ($p < .05$). Figure 4 includes graphic representations from which it can be seen that Physical Science Study Curriculum (PSSC) was superior to traditional physics (TP) and Harvard Project Physics (HPP) on the composite pretest. There was no difference between TP and HPP on the composite pretest. This same ordering was found on the composite posttest. For TP, the composite pretest mean was 46.50 with a composite posttest mean of 50.66, producing a change score of 4.16. PSSC had a composite pretest mean of 49.53 with a composite posttest mean of 53.59, a change of 4.06. The composite pretest mean was 46.62 for HPP, and the composite posttest mean was 51.86, yielding a change score of 5.24. Although each curriculum showed significant gain from pretesting to posttesting, there was no evidence of difference in composite change scores at the .05 level among the three physics courses.

Significant differences were found among the composite (combined student scores from three curriculum) scores of the physics achievement test, Dunning-Abeles Physics Achievement Test, Form E (DAPT); critical thinking test, Watson-Glaser Critical Thinking Appraisal (WGCTA); and science processes, Wisconsin Inventory of Science Processes (WISP). In Figure 5 is displayed a pretest mean of 49.56 for the WGCTA which is not statistically different from the WISP pretest mean of 48.89.

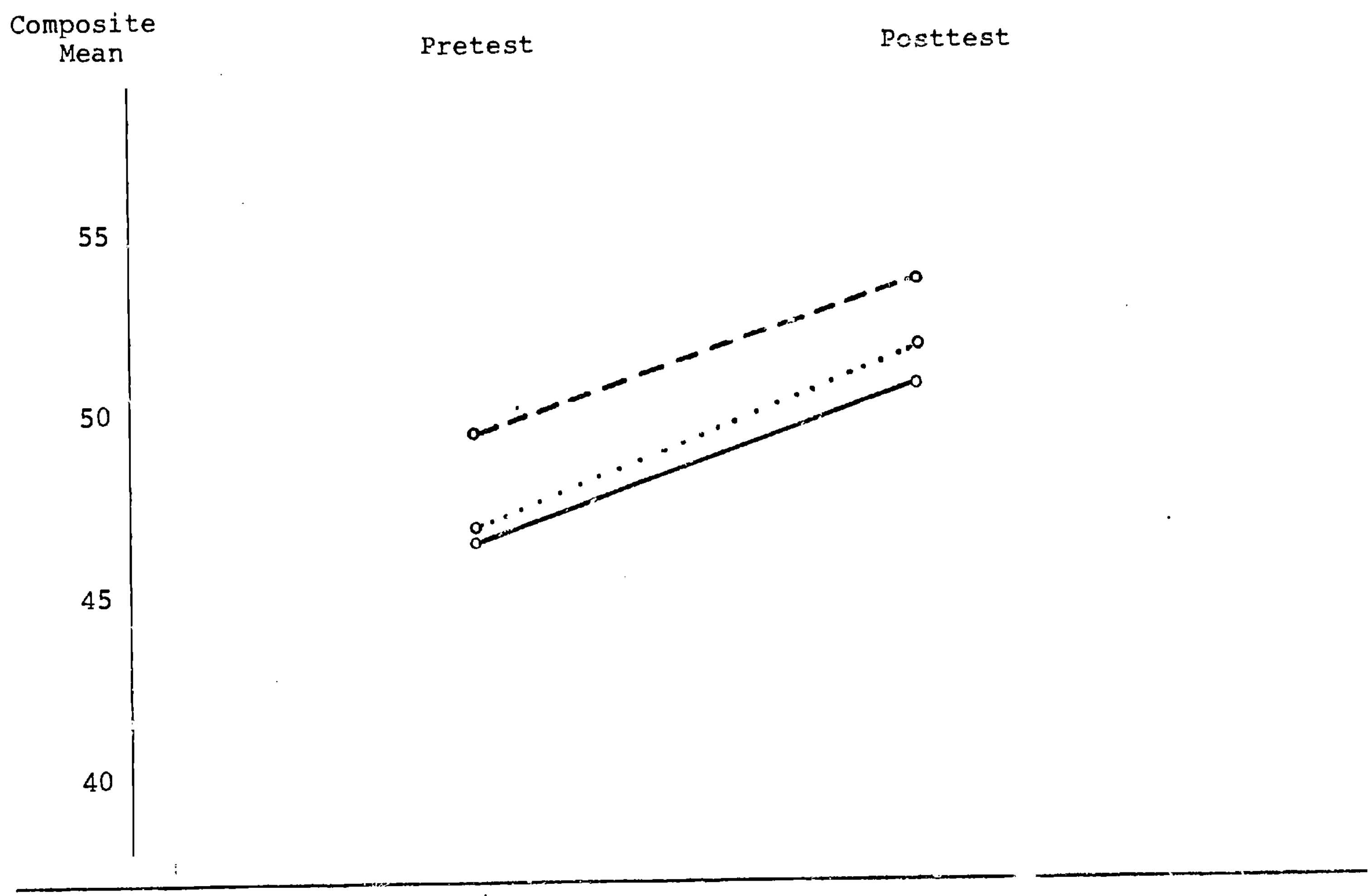

Figure 4

Comparison of Composite Test Scores Among Curricula

TP ————	Pretest 46.50	Posttest 50.66
PSSC - - - - -	Pretest 49.53	Posttest 53.59
HPP	Pretest 46.62	Posttest 51.86

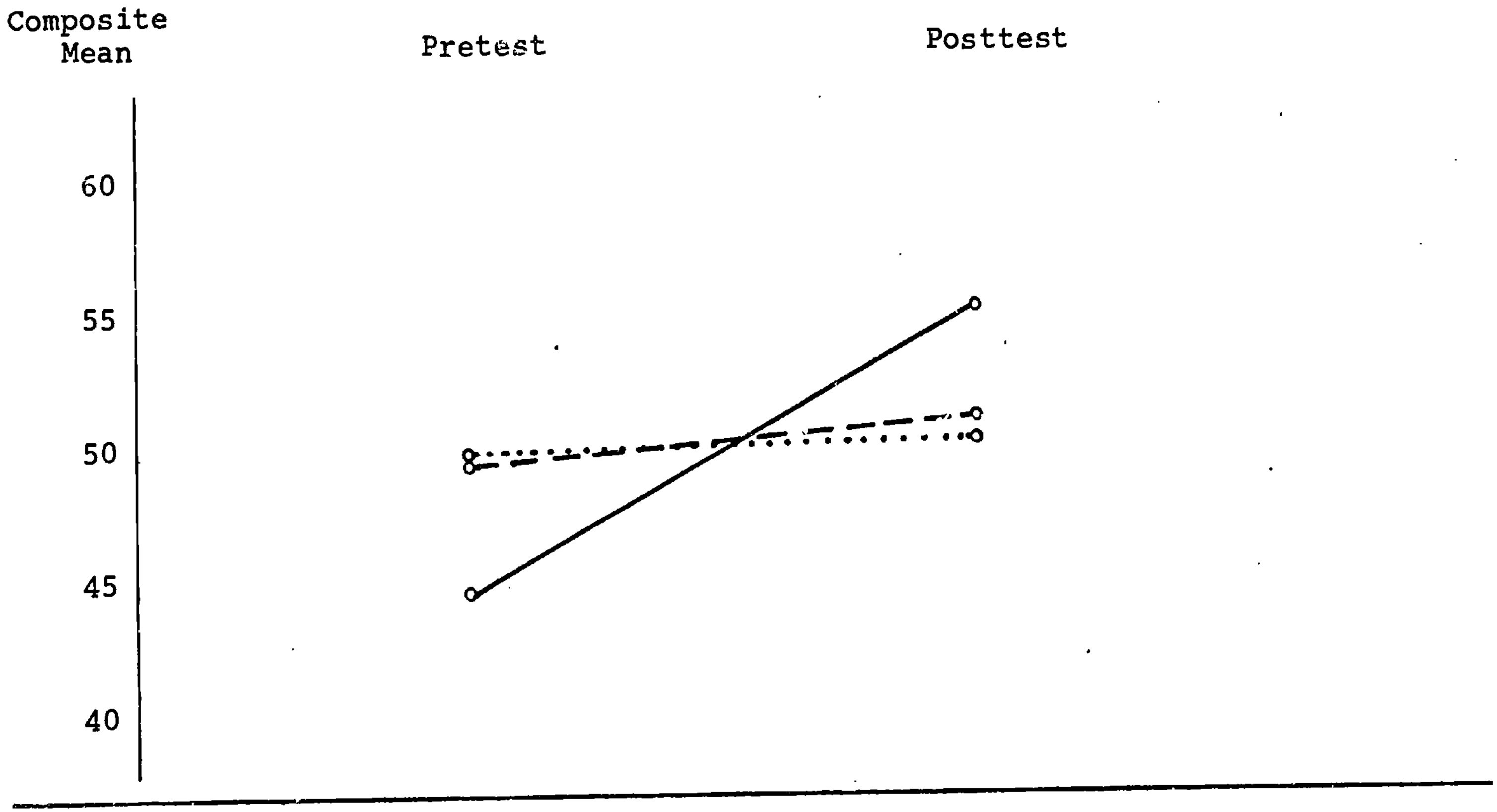

Figure 5

Comparison of Scores Among Achievement Measures

DAPT ———	Pretest 44.82	Posttest 55.18
WISP - - - -	Pretest 48.89	Posttest 51.11
WGCTA	Pretest 49.56	Posttest 50.44

The WGCTA pretest mean of 49.56 was significantly greater than the DAPT pretest mean of 44.82 and the WISP pretest mean of 48.89 was significantly greater than the pretest mean of 44.82 on the DAPT. The pretest means may then be ordered from high to low as follows: The WGCTA and WISP pretest means were equivalent and significantly larger than the DAPT pretest mean, ((WGCTA = WISP) > DAPT). A reversal of the ordering of the three means was found on posttest results. The DAPT posttest mean was significantly larger than the WGCTA and WISP posttest means, which again were not found to differ (DAPT > (WGCTA = WISP)).

The DAPT mean scores resulted in a positive change score of 10.37 (Pretest mean = 44.82, Posttest mean = 55.18). The WISP change score (2.22) was also positive (Pretest mean = 48.89, Posttest mean = 51.11). The WGCTA had a positive change score of 0.88 (Pretest mean = 49.56, Posttest mean = 50.44). There was a significant gain between pretesting and posttesting for each test at the .05 level. The DAPT change score of 10.37 was significantly greater than a change score of 2.22 on the WISP and a change score of 0.88 on the WGCTA. The change score of 2.22 on the WISP was significantly greater than the change score of 0.88 on the WGCTA. The relative changes on the tests may be ordered as follows: The DAPT change was greater than WISP change, which was in turn greater than WGCTA change, all changes being positive (DAPT change > WISP change > WGCTA change).

The interaction between curricula and individual tests was not statistically significant; this finding led to the default acceptance of the null statement of Hypothesis II.

Analysis of Classroom Verbal Behavior

Three methods of analysis were used to make comparisons of the classroom verbal interaction matrices. Darwin's Chi-square technique was utilized to compare total verbal interaction patterns of a matrix. The Kruskel-Wallis _H_-test was utilized to compare individual categories among three curricula. The two composite verbal interaction matrices (high scoring classes versus low scoring classes) were analyzed by the Mann-Whitney _U_-test to discriminate among the individual categories of the matrices.

The matrix comparisons are presented in the following sequence: comparison of the three verbal interaction matrices composed of TP, PSSC, and HPP classes and then an analysis of the two high and low scoring class verbal interaction matrices, with additional analyses of high scoring classes.

Hypothesis III was concerned with the differences among composite verbal interaction matrices for the three curricula with respect to the total verbal interaction pattern for all ten categories. This hypothesis was evaluated by analyzing the total classroom verbal behavior for the three physics curricula, based upon the formation of composite matrices from individual verbal interaction matrices. [Composite matrices of the TP, PSSC, and HPP classes, respectively, are displayed in Appendix D.] The result of applying Darwin's Chi-square technique indicated differences at the .05 level in the total verbal interaction patterns among the curricula. This finding resulted in the rejection of the null form of Hypothesis III.

Hypothesis IV dealt with the differences among the classroom verbal interaction matrices for the three curricula with regard to the percentage frequencies in each of the ten categories, and the differences among the classroom matrices for the three curricula with regard to four selected aspects of verbal interaction. The hypothesis was tested in two segments. Table 4 presents the percentage frequencies of the ten Flanders categories for the three curricula and the Kruskel-Wallis H-values for each comparison. Table 5 presents the mean ratios for the selected aspects of classroom verbal behavior among the three physics curricula and the respective Kruskel-Wallis H-values. The application of the Kruskel-Wallis H-test to both the ten categories and the four selected aspects did not yield evidence of differences among the physics curricula with respect to these variables. Therefore, the null statement of Hypothesis IV was not rejected.

Hypothesis V was concerned with the difference between the high and low scoring classroom composite verbal interaction matrices with regard to the total interaction pattern for all ten categories. The composite matrix for the ten highest scoring

TABLE 4

Comparisons of Curricula on the Percentage Frequencies in the Ten Flanders Categories

Flanders Category	Mean Percentage TP	PSSC	HPP	H-value
1	0.00	0.05	0.07	1.07
2	2.18	1.36	2.78	0.11
3	7.38	6.14	2.38	2.09
4	6.18	10.99	4.78	0.69
5	56.30	38.87	69.62	4.98
6	1.73	0.99	1.11	5.14
7	0.30	0.72	0.35	1.68
8	3.21	10.96	4.22	1.64
9	14.73	13.13	7.55	3.44
10	8.00	11.93	7.14	2.81

n.b. values for *H* must equal or surpass 5.99 to be significant at the .05 level.

TABLE 5

Comparisons of Selected Flanders Aspects of Classroom Verbal Interaction in Three Physics Curricula

Flanders Aspect	Mean Ratio TP	PSSC	HPP	H-value
Indirect-Direct Ratio	0.23	0.31	0.18	2.85
Teacher-Talk Ratio	0.73	0.57	0.78	5.25
Student-Talk Ratio	0.19	0.24	0.14	2.37
Content Ratio	0.62	0.48	0.70	3.92

n.b. values for H must equal or surpass 5.99 to be significant at the .05 level.

classes was based on composite achievement scores disregarding curriculum. The composite matrix for the ten lowest scoring classes was calculated in a parallel manner. [These composite matrices are presented in Appendix E.] Application of Darwin's Chi-square technique indicated that the two total interaction patterns were significantly different ($p < .05$). Thus, the null statement of Hypothesis V was rejected.

Hypothesis VI dealt with the differences between the high scoring and low scoring composite verbal interaction matrices with respect to the percentage frequencies in each of the ten categories, and four selected aspects of classroom verbal behavior. The mean percentage frequencies and the Mann-

Whitney U-values for the ten categories are presented in Table 6. The mean ratios and Mann-Whitney U-values for the four selected aspects are presented in Table 7. The application of the Mann-Whitney U-test did not provide evidence of differences between the high and low scoring classes for the ten categories of verbal interaction or evidence of differences between the high and low scoring classes as to the four selected aspects of classroom verbal behavior. These findings led to the default acceptance of the null statement of Hypothesis VI.

TABLE 6

Comparison of the High and Low Scoring Classes on the Percentage Frequencies in Each of the Ten Flanders Categories

Flanders Category	Mean Percentage High	Mean Percentage Low	U-value
1	0.16	0.00	40
2	1.23	1.56	44
3	9.42	3.62	30
4	7.20	4.27	42
5	57.39	52.80	48
6	0.87	1.32	33
7	0.64	0.48	47
8	5.27	3.79	46
9	11.92	12.15	42
10	6.07	19.56	28

n.b. values for U must be equal to or less than 27 to be significant at the .05 level.

TABLE 7

Comparison of Selected Flanders Aspects of Classroom Verbal Behavior for the High and Low Scoring Classes

Flanders Aspect	Mean Ratio High	Mean Ratio Low	U-value
Indirect-Direct Ratio	0.26	0.17	37
Teacher-Talk Ratio	0.71	0.64	37
Student-Talk Ratio	0.17	0.16	46
Content Ratio	0.65	0.57	41

n.b. values for the U must be equal to or less than 27 to be significant at the .05 level.

Hypothesis VII dealt with the differences among the classroom verbal interaction matrices for the high scoring classes of the three curricula with regard to the percentage frequencies in each of the ten categories and four selected aspects of verbal interaction. The mean percentage frequencies and the Kruskel-Wallis H-values for each comparison are presented in Table 8.

The mean ratios for the selected aspects of classroom verbal behavior among high scoring classes for the three physics curricula and the respective Kruskel-Wallis H-values are shown in Table 9.

Table 8

Comparisons of Curricula for High Scoring Classes on the Percentage Frequencies in the Ten Flanders Categories

Flanders Category	Mean Percentage TP	PSSC	HPP	H-value
1	0.00	0.22	0.00	0.72
2	0.43	0.99	1.34	0.61
3	17.61	5.75	6.00	1.61
4	5.99	11.69	6.13	0.00
5	44.97	55.13	73.60	2.75
6	0.92	1.08	2.55	1.37
7	0.62	0.84	0.78	0.18
8	2.10	10.55	3.47	0.86
9	13.62	8.84	2.44	4.46
10	9.39	4.70	4.78	2.75

n.b. values for H must equal or surpass 4.71 to be significant at the .05 level.

Table 9

Comparisons of Selected Aspects of Classroom Verbal Interaction Among Three Physics Curricula for High Scoring Classes

Verbal Interaction Aspect	Mean Ratio TP	PSSC	HPP	H-value
Indirect-Direct Ratio	0.33	0.42	0.14	1.20
Teacher-talk ratio	0.71	0.75	0.89	3.93
Student-talk ratio	0.20	0.18	0.06	1.46
Content ratio	0.52	0.66	0.79	2.79

n.b. values for H must equal or surpass 4.71 to be significant at the .05 level.

The application of the Kruskel-Wallis H-test to both the Flanders ten categories and the four selected aspects did not yield evidence of differences among the high scoring classes for the three curricula with respect to these variables. Therefore, the null statement of Hypothesis VII was accepted. These findings are in parallel to those of Hypothesis IV and VI, neither of which were rejected.

Summary

The present section presents the outcomes component of the evaluation model. The outcomes may be classified as student achievement (analysis of test data), classroom climate (analysis of classroom verbal behavior), and student achievement versus classroom climate (analysis of classroom verbal behavior for high and low scoring classes).

PSSC students exhibited superior performance on composite achievement measures when compared with TP and HPP students. This superior achievement was found not only on entry to the curriculum but also on completion of that curriculum. However, the relative changes in student composite achievement for the three curricula were not found to differ.

When comparing student achievement on each of the three measures of achievement, students exhibited equivalent ability as to critical thinking and understanding of science processes. However, upon entry to the curriculum, student achievement as to physics content was found to be significantly lower than the other two components of science knowledge mentioned previously, and significantly greater than the other two components at the completion of the academic year. In terms of increased ability, change in physics content achievement was greater than change in understanding science processes, which, in turn, was greater than change in critical thinking.

Student achievement on individual tests did not differ among the three curricula, indicating that no interactive effect occurred.

When comparisons were made among the three curricula for the classroom climate variables, it was found that the curricula differed for the total verbal interaction pattern. However, subsequent analyses of the ten categories and four selected aspects failed to isolate the differences in a clear and interpretable fashion. This can be interpreted to mean that evidence for differing classroom verbal behavior exists among these curricula, but the Flanders categories and selected aspect ratios, in the currently used form, are insufficient to isolate the specific sources.

The classroom climate, i.e., total verbal interaction pattern, of the classes exhibiting high student achievement was found to differ from the classroom climate of the classes exhibiting low student achievement. Analyses of the ten categories and four selected aspects of verbal behavior again failed to isolate specific variables which contributed to the difference found in the total verbal interaction patterns, most likely for the reasons cited previously.

Judgments, Discussion, and Recommendations

Science educators have expressed the need for evaluation of the modern science curricula which have been developed in the past ten years. In response to this need a five component evaluation model was developed to aid in systematizing the evaluation process and to insure the consideration of key variables in such evaluation processes.

The evaluation model was applied to a specific area of science education, physics, and its three major curricula--traditional physics (TP), Physical Science Study Curriculum (PSSC), and Harvard Project Physics (HPP)--as currently used in American high schools. The sample employed in the application of the model included 24 secondary schools in 7 states with 38 classes containing 954 physics students. Due to the random procedures used in sample selection, the sample may be considered as representative of high school physics students and curricula in the southwestern, midwestern, and western areas of the United States.

The behavioral goals, the first component of the model, consisted of two basic objectives of physics, mastery of physics content, and acquisition of the process of science. The formative experiences, the second component of the evaluation model, consisted of assessing length of time in the educational environment, academic aptitude, and entry level of attainment in both physics content and science processes of the students in the physics curricula. Sex of the student was also considered as a potential element of cultural role differentiation in performance. The third component, transactions, consisted of the teaching triad. The fourth component, outcomes, consisted of the metrics of classroom climate--student cognitive behavior and the classroom verbal behavior. This classroom verbal behavior was quantified from audio tape recordings of both students and teachers involved in the three curricula, the recordings being made at several different times during the academic year. The judgments component of the model, the fifth and last, consisted of the subsequent analyses of quantifications leading to decision making as to relationships exhibited among teachers, students, and physics curricula.

Although background characteristics of age, sex, academic aptitude, and prior education may generally be considered as pertinent factors influencing student achievement (Thurstone, 1947; Ausubel & Robinson, 1969), the results of formative experiences for the three groups of students (TP, PSSC, and HPP) were found to function homogeneously with respect to physics achievement. Results of this study indicate that high school physics students are of above average academic ability and that neither grade level nor sex of a student is a determining factor as to the level of achievement within any physics curriculum.

The students within each of the three curricula displayed significant increase in physics achievement, not only in terms of the composite measures of physics achievement but also in respect to each of the three tested elements of physics achievement--physics content, science processes, and critical thinking. This increase in physics achievement was found typical of all curricula and therefore no ordering of curricula upon a continuum of "effectiveness" was possible. There was no evidence of a specific curriculum interacting in a differential manner with a particular component of physics achievement. On the bases of the student achievement measures, the three curricula must be judged as effective, as previously defined, and equivalent in implementation of learning of physics.

Student achievement was found to differ as to particular aspects of physics achievement in that the increase in physics content mastery was greater than the increase in understanding science processes. Although the students' critical thinking ability was found to have increased significantly, this change was significantly less than the change exhibited in the other two aspects of physics achievement. When viewed through the evaluation model, these findings led to the judgment that all three physics curricula were more effective in producing learning of physics content than in developing an understanding of science processes, and least effective in developing critical thinking. This ordering of the three components as to effectiveness must be viewed within the framework that the curricula were found effective in each component and the ordering is relative to magnitude of change in final functioning level of students.

The three curricula were found to differ as to classroom verbal interaction patterns; however, due to the nature of the Flanders' system for quantification, analyses permissible for isolating causal variables failed to provide specific evidence as to the sources of differences within patterns. The Flanders' system for quantification provides data of nominal scaling and therefore limits statistical analyses to nonparametric tests which may be considered less robust than parametric tests. Within the evaluation model context, this finding led to the judgment that the three teacher-student-curriculum triads were different but clarification as to the exact nature of the differences was not possible.

Although statistical analyses of the classroom verbal interaction data did not reveal or isolate the specific differences among the total verbal interaction patterns for the three curricula, there are certain trends that appear in the data that may explain differences among the classroom verbal interaction patterns for the three triads.

The TP triad represents a classroom climate in which lecture is the principal mode of verbal behavior but to a lesser extent than HPP and to a greater extent than PSSC. Of the three curricula, the TP students initiated more talk in class and the teachers expanded on the student ideas more than did the teachers within the other curricula. TP teachers tend to ask very few convergent questions when compared to PSSC teachers. Although the teacher-talk ratio (.73) is relatively high when compared to the student talk ratio (.19), there is considerable freedom for the student to initiate questions and a tendency for the teacher to expand on these ideas. Although the TP teacher is quite direct and concentrates on content dissemination, there is a freedom for students to interact with the teacher and for their ideas to be utilized by the teacher.

In the PSSC triad, there exists a tendency for the classroom to be less lecture controlled than either TP or HPP; however, there are many convergent questions asked by the teacher. The TP triad and PSSC triad share common patterns of verbal behavior in teachers' accepting the building onto student ideas. Students in PSSC, as in TP, appear to feel free to initiate questions and the teacher spends time in developing the ideas suggested. The teacher-talk ratio (.56) and student-talk ratio (.24) are the smallest and largest respectively of the three triads.

The most surprising results of trends in classroom verbal interaction patterns were exhibited in the HPP triad. HPP teachers demonstrated the highest frequency of lecture to the classroom for the three groups. Nearly 70% of the classroom verbal activities were devoted to lecture as compared to 56% for TP and 39% for PSSC. There was very little acceptance of student ideas and relatively few student initiated questions. HPP teachers demonstrated a reluctance to solicit student statements either by direct questioning or by providing a climate for the students to initiate ideas of their own. The teacher-talk ratio (.78) and student-talk ratio (.14) further exhibited the large frequency of teacher directed verbal activities when compared to verbal activities of the HPP students. This may indicate that verbal interaction in HPP classes was not the type expected in a curriculum that has less structure for the student in its design. This may be a limiting factor for the HPP student.

When summarizing the verbal interaction patterns for the three triads there are some common characteristics of the verbal classroom climate that appear. In each triad, the physics teachers were quite conservative with praise or encouragement for the students although the teachers did not criticize or give lengthy directions. The primary concern of the teachers as a group was the dissemination of content information.

The key differences in trends appear to focus on the type of questions asked, how frequently the students ask questions, and how extensively the teacher utilizes and expands on the students' ideas. It appears that both TP and PSSC teachers create an atmosphere in which students feel free to ask or initiate questions whereas this freedom seemed restricted in the HPP classrooms. This was not the type of atmosphere expected in the HPP curriculum and may be due, in part, to the teacher not being as secure in the course as in a course they had taught for a longer period of time. It could be due to teachers' practices, carried over from experience in physics courses taught prior to teaching HPP. Both PSSC and TP teachers spend more of the class period developing or using ideas of the students than do HPP teachers. Although the PSSC teachers asked more questions than TP or HPP teachers, these questions were convergent.

The transaction component of the evaluation model may therefore be viewed as consisting of three distinct binary relationships of classroom verbal behavior intersecting with a physics curricula. However, each distinct relationship has resulted in similar student behavior on achievement measures.

The classroom verbal interaction patterns of the classroom exhibiting high student achievement were found to differ from the patterns of classrooms with low student achievement.

Once again statistical analyses of the verbal interaction data did not reveal or isolate the specific differences between the classroom verbal interaction patterns for the highest and lowest scoring classes. There are, however, some trends that appear for high versus low scoring classes within the sample. Teachers of both groups utilized more than one-half of the class period in lecture. Very little praise or encouragement was offered the students in either group; however, the students were not given lengthy directions or criticized for their actions in the classroom. Both groups appear to have the freedom to initiate questions. Student-talk and teacher-talk ratios for both groups were similar. The key differences in trends appear in the relatively high frequency of teacher questions, usage and expansion of student ideas, and the relatively small amount of the class devoted to confusion or silence in the high scoring classes when compared to the low scoring classes.

The conclusions of the present study indicate that the evaluation model was applicable when applied to three physics curricula. This finding tends to support the contention that the major goals of physics instruction consist of student mastery of physics content and understanding science processes. Only one judgment may be considered in conflict with previous evaluations of physics curricula. Day (1964), Rutledge (1965), and Brakken (1965) reported that students enrolled in PSSC achieved at a significantly higher level when compared to students in traditional physics in terms of critical thinking ability as measured by the Watson-Glaser Critical Thinking Appraisal. The present study did support the general contention that critical thinking can be enhanced during the duration of an academic year but this increase could not be attributed solely to a particular curriculum. This may be evidence of physics instruction in TP courses evolving and being influenced in this evaluation by the more modern curricula of HPP and PSSC. Hence, reducing differences that reportedly existed earlier.

The transactions component of the model facilitated the investigation of the hypothesized relationship among teacher, student, and curriculum. The three teaching triads, differentiated by curriculum, were found to differ as to classroom climate, i.e., classroom verbal interaction patterns. The trends within the classroom verbal interaction patterns for the three triads seem to focus on the type of questions asked by the teacher, as determined indirectly from student answers, how frequently the students ask questions, and how extensively the teacher expands on the student's ideas. It appears that both TP and PSSC teachers create a learning environment in which students feel free to ask or initiate questions whereas this freedom seems restricted in the HPP classrooms. Both TP and PSSC teachers spent more of the class period developing or using ideas of students than HPP teachers. Although the PSSC teachers asked more questions than TP or HPP teachers, these questions were convergent.

These trends suggest the possibility that TP and PSSC teachers were more familiar with their curricular materials than HPP teachers since it was the first year HPP had been utilized in the high schools. Since the previous comparative studies have not been conducted among the three curricula as to this relationship, there is need for additional investigation to clarify and further describe the relationships of curricula to classroom verbal interaction patterns. There is also critical need for either more refined measures of classroom verbal interaction or for statistical techniques which are more robust in their ability to deal with nominally scaled data.

A similar condition exists with respect to the indicated difference in classroom verbal behavior between high and low achieving classes. The major differences in trends of the verbal interaction patterns appear in the relatively high frequency of teacher questions, usage and expansion of student ideas, and the relatively small amount of class devoted to silence or confusion in the high scoring classes as compared to the low scoring classes. Of the reviewed studies, only one has been conducted in this area of physics education; Pankratz (1966), although employing different criteria of effectiveness, reported results similar to those reported here.

The findings of this study suggest the necessity for further research in the area of physics education. The _transactions_ component of the model has identified certain trends that appear in the classroom verbal behavior patterns within the three triads, and also trends that appear in the high and low scoring classes. These trends suggest that possibly other factors, not considered in this study, might be worthy of consideration in future studies. Future studies might include the students' attitude toward teacher and course and the teachers' attitude toward the course and how these attitudes effect student achievement.

It is further recommended that within each curriculum, research be conducted in structured versus unstructured classrooms. There was some evidence in this study which indicated that students in unstructured classrooms achieved to a lesser extent than those in a more structured environment. It may be that student outcomes measured in this study correlate higher with a structured classroom situation than with unstructured and that outcomes of unstructured classrooms are different, to some extent, than those of structured classrooms.

One of the most impressive findings of this study was the degree to which the physics teachers--in totem, exhibited serious-minded, concentrated, and dedicated efforts to enhance the learning of each and every student. In observing the classrooms, a panorama of teaching techniques was noted. This panorama included student centered activities where the teacher was a resource person in the purest sense, multi-media presentations, small group discussions and investigations, and independent study by the students. Perhaps this high ability level of the teachers, irrespective of curriculum, is the primary factor contributing to the impressive effectiveness of each curriculum in enhancing student achievement.

The HPP teachers, all of whom participated in formal instruction to familiarize them with the HPP curriculum prior to implementation, were successful in utilizing some of the various teaching methods provided by the curriculum. The teachers utilized activities involving student reports, individual laboratory investigations, and lecture-discussions in implementing the curriculum with ninth grade students as well as twelfth grade students. Although the design of the course provides for maximum flexibility with regard to teaching strategies, the teachers did not demonstrate flexibility in classroom verbal interaction.

The PSSC teachers, utilizing relatively few basic concepts of physics, a multitude of audio-visual materials and inquiry methods, were able to interact with their eleventh and twelfth grade students in such a manner as to apparently fulfill the objectives of the curriculum.

The TP teachers were surprisingly verbal inquiry oriented in methodology. Apparently, they have taken the classical physics units of the past and brought new life to the material. The broad spectrum of modern teaching methodology was interfaced with tradition in their classrooms resulting in a physics curriculum which can not be considered "traditional." The high degree of effectiveness of such an interface is exhibited by the fact that ten of the eleven TP classes achieved outstandingly on each aspect of physics knowledge.

The findings and observations tend to suggest that the crucial factor determining effectiveness of a physics curriculum is the expertise of the teacher. This expertise must not only be in physics content, but also in implementation of the scientific methods of psychology and learning theory. Therefore, based upon this study, the recommendation must be made that preparatory and in-service teacher training programs emphasize flexibility in providing for individual differences and establishment of a conducive socio-emotional climate for the classroom.

The limitation in applying the evaluation model appears to be the lack of precision in quantifying and analyzing data from the <u>transactions</u> component. The recommendation must be made that due to the crucial need for information in this area of evaluative models, intensive refinement of measuring instruments is of utmost importance. It has been demonstrated that the proposed evaluative model is applicable and produces judgments that can lead to effective curriculum comparisons. The applicability of the present model is limited by the quality of data available from the classroom behavior instruments and the techniques for their analysis. The development of stronger uses of the proposed model--or others which might parallel its utility--appear to depend upon strengthening the definition and assessment of the elements that make up the complex interplay of teachers, students, curricular materials, and instructional techniques in the classrooms.

BIBLIOGRAPHY

Bibliography

Ackerson, P. B. A study of the relationship between achievement in PSSC physics and experience in recently developed sources in science and mathematics, _Dissertation Abstracts_, 27:1:44-A, July, 1966.

Amidon, E. J., & Flanders, N. A. _The role of the teacher in the classroom_. Minneapolis: Amidon and Associates, 1963.

Anderson, J. S. A comparative study of chemical education material study and traditional chemistry in terms of cognitive processes, _Dissertation Abstracts_, 25:9:5147-5148, March, 1965.

Anderson, L. H. The measurement of domination and of socially integrative behavior in teachers' contacts with children. _Child Development_, 1939, 10:73-89.

Aschner, Mary Jane McCue. The analysis of verbal interaction in the classroom. In A. A. Bellack (Ed.), _Theory and research in teaching_. New York: Bureau of Publications, Teachers College, Columbia University, 1963.

________. The language of teaching. In B. Othaniel Smith and Robert H. Ennis (Eds.), _Language and concepts in education_. Chicago: Rand McNally & Company, 1961.

Atwood, Ronald K. CHEM study achievement among groups classified by cognitive preference scores. _Journal of Research in Science Teaching_, 1967, 5:154-59.

Ausubel, D. P., & Robinson, F. G. _School learning: an introduction to educational psychology_. New York: Holt, Rinehart and Winston, Inc., 1969.

Beaucamp, Wilbur. _Instruction in science_. (U.S. Office of Education Bulletin, 1932, No. 17, Monograph No. 22). Washington: Government Printing Office, 1933.

Belanger, Maurice. Learning studies in science education. _Review of Educational Research_, 1969, 39:4:377-95.

________. The study of teaching and the new science curricula. _The Science Teacher_, 1964, 31:7:31-35.

Bellack, Arno A., Keibard, Herbert M., Hyman, Ronald T., & Smith, Frank L., Jr. _The language of the classroom_. New York: Teachers College, Columbia University Press, 1966.

Berry, W. E. The comparative effects of PSSC physics and traditional physics on student achievement, Dissertation Abstracts, 27:4:878-A-879-A, October, 1966.

Biddle, Bruce J. The integration of teacher effectiveness Research. In Bruce J. Biddle and William J. Ellena (Eds.), Contemporary research on teacher effectiveness. New York: Holt, Rinehart and Winston, Inc., 1964.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain. New York: David McKay Company, Inc., 1956.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of varianc problems. Annuals of Mathematical Statistics, 1954, 25:290-302.

Brakken, Earl. Intellectual factors in PSSC and conventional high school physics. Journal of Research in Science Teaching, 1965, 3:19-25.

Brown, H. E. Water-tight compartments. School Science and Mathematics, 1939, 39:840-45.

Champlin, R. F., & Hassard, J. R. A comparative study of two earth science courses. Unpublished Master's thesis, Boston University, 1966.

Citron, Irvin M., & Barnes, Cyrus W. The search for more effective methods of teaching high-school biology to slow learners through interaction analysis: Part I: the effects of varying teaching patterns. Journal of Research in Science Teaching, 1970, 7:9-19.

Cooley, William W., & Klopfer, Leo. Manual for the test on understanding science. Princeton: Educational Testing Service, 1961.

Coulton, John C. The effectiveness of inductive laborary, inductive demonstration, and deductive laboratory in biology. Journal of Research in Science Teaching, 1966, 4:185-86.

Craven, Gene F. Critical thinking abilities and understanding of science by science teacher-candidates at Oregon State University. Dissertation Abstracts, 27:1:125-A, July, 1966.

Cronbach, L. Evaluation for course improvement. Teachers College Record, 1963, 64:672-83.

Crumb, G. H. A study of understanding science developed in high school physics, <u>Dissertation Abstracts</u>, 26:3:1506-07, September, 1965.

———. Understanding of science in high school physics. <u>Journal of Research in Science Teaching</u>, 1965, 3:246-50.

Darwin, J. H. Note on the comparison of several realizations of a Markoff chain. <u>Biometrika</u>, 1959, 46:412-19.

Day, W. F. An experimental evaluation of PSSC and traditional physics in six areas of critical thinking. <u>American Journal of Physics</u>, 1964, 32:11:901.

Dunning, Gordon M., & Abeles, Sigmund. <u>Dunning-Abeles physics test, form e</u>. New York: Harcourt, Brace & World, Inc., 1967.

Evans, Thomas P. An exploratory study of the verbal and non-verbal behaviors of biology teachers and their relationships to selected personality traits. <u>Dissertation Abstracts</u>, 29:5:1359-A-1360-A, November, 1968.

———, & Balzer, LeVon. An inductive approach to the study of biology teacher behaviors. <u>Journal of Research in Science Teaching</u>, 1970, 7:47-56.

Flanders, N. A. <u>Teacher influence, pupil attitudes, and achievement</u>. (U.S. Department of Health, Education and Welfare, Office of Education, Cooperative Research Project No. 397.) Minneapolis: University of Minnesota, 1960.

———. Teacher influence, pupil attitude and achievement. In <u>Cooperative Research Monograph</u>. (U.S. Department of Health, Education, and Welfare, Office of Education, Cooperative Research Monograph No. 12, OE 25040.) Washington: Government Printing Office, 1965.

———, & Amidon, E. J. <u>The role of the teacher in the classroom: a manual for understanding and improving teachers' classroom behavior</u>. Minneapolis: Association for Productive Teaching, Inc., 1967.

Furst, Norma. The effects of training in interaction analysis on the behavior of student teachers in secondary schools. Paper read at the annual meeting of the American Educational Research Association, Chicago, February, 1965.

Gagné, R. M. Curriculum research and the promotion of learning. In Robert E. Stake (Ed.), Perspectives of curriculum evaluation. (AERA Monograph Series on Curriculum Evaluation, No. 1.) Chicago: Rand McNally & Company, 1967.

Gennaro, Eugene D. A comparative study of two methods of teaching high school biology: the BSCS yellow version and labotary blocks with collateral reading. Dissertation Abstracts, 25:7:3996-3997, January, 1965.

George, Kenneth D. The effect of BSCS and conventional biology on critical thinking. Journal of Research in Science Teaching, 1965, 3:293-299.

________. An experimental evaluation of BSCS and conventional biology by comparing the effects on critical thinking ability. Dissertation Abstracts, 26:2:792-793, August, 1965.

Glaser, R. (Ed.). Teaching machines and programmed learning, II. Washington: National Education Association, 1965.

Gold, L. L. Verbal interaction patterns in the classrooms of selected science teachers: biology. Dissertation Abstracts, 27:5:1281-A, November, 1966.

Heath, R. W. Curriculum, cognition, and educational measurement. Educational and Psychological Measurement, 1964, 24:2:239-53.

________, & Stickel, David W. CHEM and CBA effects on achievement in chemistry. Science Teacher, 1963, 30:45-46.

Henkel, E. Thomas. The effects of instruction in physics upon the critical thinking ability of undergraduate students. Journal of Research in Science Teaching, 1967, 5:89-94.

Henry, Nelson B. (Ed.) The content and methods of senior high school science. In Science education in American schools. (Forty-sixth Yearbook, Part I of the National Society for the Study of Education.) Chicago: The University of Chicago Press, 1947.

Herron, J. Dudley. Evaluation and the new curricula. Journal of Research in Science Teaching, 1966, 4:159-70.

________. A factor analytic and statistical comparison of CHEM study and conventional chemistry in terms of their development of cognitive abilities. Dissertation Abstracts, 26:8:4333, February, 1966.

Holton, Gerald. Project physics. A report on its aims and current status. The Physics Teacher, 1967, 5:5:198-211.

Hough, John B. An observational system for the analysis of classroom instruction. In Edmund J. Amidon and John B. Hough (Eds.), Interaction analysis: theory, research, and application. Reading, Mass.: Addison-Wesley Publishing Company, 1967.

Hughes, M. M., & others. The assessment of the quality of teaching: a research report. (U.S. Office of Education Cooperative Research Project, No. 353.) Salt Lake City: University of Utah, 1959.

Hurd, Paul DeH. Science education for changing times. In Nelson B. Henry (Ed.), Rethinking science education. (Fifty-ninth Yearbook of the National Society for the Study of Education.) Chicago: The University of Chicago Press, 1960.

Hyman, Ronald T. Teaching: triadic and dynamic. Educational Forum, 1967, 32:65-69.

________. Ways of teaching. New York: J. B. Lippincott Company, 1970.

Karle, I. F. The effectiveness of open-ended chemistry experiments in a high school setting: a comparison of open-ended chemistry experiments with the conventional laboratory exercises in selected high school chemistry classes. Dissertation Abstracts, 21:5:1099-1100, November, 1960.

Kastrinos, William. The relationship of two methods of teaching to the development of critical thinking by high school students in advanced biology. Science Education, 1964, 48: 187-95.

Kelley, T. L. The selection of upper and lower groups for the validation of test items. Journal of Educational Psychology, 1939, 30:17-24.

Lance, Mary L. A comparison of gains in achievement made by students of BSCS high school biology and students of a conventional course in biology. Dissertation Abstracts, 25:5: 2814-2815, November, 1964.

Lee, E. W. A study of the effects of two methods of teaching high school chemistry upon critical thinking abilities. Dissertation Abstracts, 25:8:4578-4579, February, 1965.

Levin, Harry A. New perspectives on teacher competence research. Harvard Educational Review, 1954, 24:98-105.

Lewin, Kurt, Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created "social climates." Journal of Social Psychology, 1939, 10:271-299.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton-Mifflin Company, 1953.

Lisonbee, L. K. The comparative effect of BSCS and traditional biology upon student achievement. Dissertation Abstracts, 24:8:3153, February, 1964.

Lisonbee, Lorenzo. The comparative effect of BSCS and traditional biology on student achievement. School Science and Mathematics, 1964, 64:594-98.

Lockard, J. David (Ed.). Sixth report of the international clearinghouse on science and mathematics curricular developments. College Park: University of Maryland, 1968.

Mager, R. F. Preparing instructional objectives. Palo Alto: Fearon Publisher, 1962.

Marks, Ronald L. CBA high school chemistry and concept formation. Journal of Chemical Education, 1967, 44:471-74.

National Education Association. Ten criticisms of public education. Research Bulletin, 1957, 35:161.

National Science Teachers Association, Committee on Issues. NSTA positions on critical issues confronting the science teaching profession. The Science Teacher, 1970, 37:7:55-56.

Newport, John F., & McNeill, Keith. A comparison of teacher-pupil verbal behaviors evoked by science--a process approach and by textbooks. Journal of Research in Science Teaching, 1970, 7:191-95.

Niman, John. Mathematical models of physics for teaching. Journal of Research in Science Teaching, 1970, 7:29-30.

Pankratz, R. S. Verbal interaction patterns in the classrooms of selected science teachers: physics. Dissertation Abstracts, 27:5:1296-A, November, 1966.

Parakh, Jal S. A study of teacher-pupil interaction in high school biology classes. Part II. Description and analysis. Journal of Research in Science Teaching, 1968, 5:183-92.

Perkins, H. V. Classroom behavior and under-achievement. American Educational Research Journal, 1965, 2:1-12.

Popham, W. J. (Ed.). Instructional objectives. Chicago: Rand McNally & Company, 1969.

Powell, E. R. Some relationships between classroom process and pupil achievement in the elementary school. Unpublished doctoral dissertation, Temple University, 1968.

Psychological Corporation. Earth science curriculum project education program end-of-year report, 1964-1965. New York: Psychological Corporation, 1965.

Rainey, Robert G. The effects of directed versus non-directed laboratory work on high school chemistry achievement. Journal of Research in Science Teaching, 1965, 3:286-92.

Rickert, Russell K. Developing critical thinking. Science Education, 1967, 51:24-27.

Rutherford, F. James. Flexibility and variety in physics. The Physics Teacher, 1967, 5:5:215-21.

Rutledge, J. A. Physics and critical thinking: an experimental evaluation of PSSC and traditional physics in six areas of critical thinking while controlling for intelligence, achievement, course background, and mobility by analysis of covariance. Dissertation Abstracts, 25:7:4197, January, 1965.

Sawyer, Robert L. An investigation of the effectiveness of the program recommended by the physical science study committee. Dissertation Abstracts, 24:12:1:5254-5255, June, 1964.

Schirner, S. W. A comparison of student outcomes in various earth science courses taught by seventeen Iowa teachers. Unpublished doctoral dissertation, University of Iowa, 1967.

Scott, W. A. Reliability of content analysis: the case of nominal coding. Public Opinion Quarterly, 1955, 19:3:321-25.

Siegel, Sydney. Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company, Inc., 1956.

Smith, B. Othaniel. A concept of teaching. Teachers College Record, 1960, 61:5:229-41.

Smith, B. Othaniel. On the anatomy of teaching. Journal of Teacher Education, 1956, 7:339-46.

_______, & others. A study of the logic of teaching: a report on the first phase of a five-year research project. (U.S. Office of Education Cooperative Research Project, No. 258 (7257)). Urbana: University 9f Illinois, 1960.

Snider, R. M. A project to study the nature of physics teaching using the Flanders' method of interaction analysis, Dissertation Abstracts, 26:12:1:7183-7184, June, 1966.

Soar, R. An integrative approach to classroom learning. (Public Health Service, Final Report, No. 7-R11MH02045.) Philadelphia: Temple University, 1966.

Stake, Robert E. The countenance of educational evaluation. In Richard C. Anderson, Gerald W. Faust, Marianne C. Roderick, Donald J. Cunningham, and Thomas Andre (Eds.), Current Research on Instruction. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969.

Stollberg, Robert. The status of science-teaching in elementary and secondary schools. In Nelson B. Henry (Ed.), Rethinking Science Education. (Fifty-ninth Yearbook of the National Society for the Study of Education.) Chicago: The University of Chicago Press, 1960.

Thomas, Barbara S. An analysis of the effects of instructional methods upon selected outcomes of instruction in an interdisciplinary science unit. Dissertation Abstracts, 29:6: 1830-A-1831-A, December, 1968.

Thomas, Dorothy A., and others. Some new techniques for studying social behavior. New York: Columbia University Press, 1929.

Thurstone, L. L. Multiple-factor analysis: a development and expansion of the vectors of mind. Chicago: The University of Chicago, 1947.

Trent, John. The attainment of the concept "understanding science" using contrasting physics courses. Journal of research in Science Teaching, 1965, 3:224-29.

Troxel, Verne A. Analysis of instructional outcomes of students involved with three courses in high school chemistry. Dissertation Abstracts, 29:6:1832-A, December, 1968.

Tyler, R. W. Basic principles of curriculum and instruction. Chicago: The University of Chicago Press, 1950.

Van Dalen, D. B. Relationship of fact and theory in research, Educational Administration and Supervision, 1959, 45:271-74.

Victor, Edward, & Lerner, Marjorie S. Readings in science education for the elementary school. New York: The Macmillan Company, 1967.

Watson, Fletcher F. Why do we need more physics courses? The Physics Teacher, 1967, 5:5:212-13.

Watson, Goodwin, & Glaser, Edward M. Watson-Glaser critical thinking appraisal manual for forms ym and zm. New York: Harcourt, Brace & World, Inc., 1964.

Welch, Wayne W. Curriculum evaluation. Review of Educational Research, 1969, 39:4:429-43.

________, & Pella, Milton O. The development of an instrument for inventorying knowledge of the processes of science. Journal of Research in Science Teaching, 1968, 5:64-68.

Welch, Wayne W., & Rothman, Arthur I. The success of recruited students in a new physics course. Science Education, 1968, 52:270-73.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, Inc., 1962.

Withall, John. The development of a climate index. Journal of Educational Research, 1951, 45:93-99.

________. The development of a technique for the measurement of social-emotional climate in classrooms. Journal of Experimental Education, 1949, 17:347-61.

Yoesting, Clarence, & Renner, John W. Is critical thinking an outcome of a college physical science course? School Science and Mathematics, 1969, 69:199-206.

APPENDIXES

APPENDIX A: Correspondence with Prospective Schools

UNIVERSITY OF HOUSTON
COLLEGE OF EDUCATION
CULLEN BOULEVARD
HOUSTON, TEXAS 77004

CULUM AND INSTRUCTION
SECONDARY EDUCATION

July 22, 1968

Dear Colleague,

The strength of the nation's science programs has grown rapidly in the past few years. This improvement is partially attributed to teachers who have willingly taken an active part in the various supportive research projects. You are invited to participate in a study of national interest in physics education at the University of Houston, Houston, Texas.

The study itself will involve Harvard Project Physics, PSSC, and traditional physics. We are conducting research which centers around processes and outcomes in the various physics courses. The evaluation techniques will consist of pretesting in September, 1968, and posttesting in early May, 1969. Testing time will be approximately two and one-half hours each time. During the year we will tape (audio) approximately three classes. This taping would cause minimal disturbance in your normal classroom procedures.

We feel the findings of this study will serve the interests of physics education on the secondary level and should prove helpful to you in evaluating your physics program. All scores and interpretations will be available to you. We will furnish you with individual scores of each student on both the pre- and posttests. These scores will be on standardized instruments so you could compare with established norms. Furthermore, we will furnish you with the mean scores from other schools participating in the study. However, no mention of schools will be made. If you so desire, the test scores could be used for evaluating individual student progress.

The audio tapes of your class will simply be used to determine what classroom processes are being used in the various types of physics courses. The days of your taping sessions will be planned at your convenience. The tapes and their analyses will be discussed with you.

There will be no cost to any school. You will be provided with all test results, yours as well as other participating schools, and audiotape analysis cost free.

The results of our research are intended for the advancement of physics education, and this is not an attempt to <u>evaluate an individual school, teacher, or student</u>. All information will be treated as confidential data. The results of our study will be made available to you in summary form again at no cost to you or your school.

Without the help of interested and concerned teachers in the field of physics, our study cannot be successfully accomplished. This makes you a very important part of this study.

If you feel that you would consider participating in such a study, would you fill out the following questionnaire.

	YES	NO	
1.	☐	☐	Would you be willing to participate in this research if your administration gives its approval?

2. Name ____________________ Home Address ____________________ (Street)

   City ______________, State ______________ Phone ______________

3. Which of the following physics courses do you plan to teach during the 1968-1969 school year?

   ☐ PSSC ☐ Traditional Physics

   ☐ Project Physics ☐ Other Explain ______________

4. Give the name of the school administrator we should contact in seeking permission to do research in your physics class.

   Name ____________________ School ____________________

   Address ____________________ (Street) ______________ (City), ______________ (State)

   Phone ____________________

On the basis of your response, I will contact your administration in regard to possible participation in the research.

Thank you for your cooperation and help. We look forward to working with you this fall.

Respectfully,

T. C. Smith, Jr.

T. C. Smith
Principal Investigator

Silas W. Schirner

Silas W. Schirner, Project Director
Science Education
University of Houston
Houston, Texas

APPENDIX B: Participating Schools

Participating Schools

Name of School	Location
Traditional Physics	
1. Robert E. Lee High School	Baytown, Texas
2. J. Frank Dobie High School	Pasadena, Texas
3. Brenham High School	Brenham, Texas
4. Needville High School	Needville, Texas
5. Angleton High School	Angleton, Texas
6. Booker T. Washington High School	Houston, Texas
7. Ross Sterling High School	Baytown, Texas
8. St. Pius X High School	Houston, Texas
PSSC Physics	
1. Thomas Jefferson High School	Port Arthur, Texas
2. Sam Houston High School	San Antonio, Texas
3. Kincaid School	Houston, Texas
4. St. Thomas High School	Houston, Texas
5. Mercy Academy	New Orleans, Louisiana
6. Pueblo High School	Tucson, Arizona
7. West Jefferson High School	New Orleans, Louisiana
8. Mount Carmel High School	New Orleans, Louisiana
9. Strake Jesuit School	Houston, Texas
10. San Diequito High School	Cardiff, California
Harvard Project Physics	
1. Chaminade School	St. Louis, Missouri
2. Emerson High School	Gary, Indiana
3. San Diequito High School	Cardiff, California
4. Unified High School	Belle Plaine, Kansas
5. Mercy Academy	New Orleans, Louisiana
6. Griffin High School	Springfield, Illinois
7. Aviation High School	Redondo Beach, California
8. Alhambra High School	Phoenix, Arizona

APPENDIX C: Categories and Illustrative Matrix for Flanders' Verbal Interaction Analysis

Teacher-talk Indirect Influence	1. ACCEPTS FEELING: accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
	2. PRAISES OR ENCOURAGES: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying "um hm?" or "go on" are included.
	3. ACCEPTS OR USES IDEAS OF STUDENT: clarifying, building, or developing ideas suggested by a student. As a teacher brings more of his own ideas into play, shift to category five.
	4. ASKS QUESTIONS: asking a question about content or procedure with the intent that a student answer.
Teacher-talk Direct Influence	5. LECTURING: giving facts or opinions about content or procedure; expressing his own ideas, asking theoretical questions.
	6. GIVING DIRECTIONS: directions, commands, or orders to which a student is expected to comply.
	7. CRITICIZING OR JUSTIFYING AUTHORITY: statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.
Student-talk	8. STUDENT TALK--RESPONSE: a student makes a predictable response to teacher. Teacher initiates the contact or solicit student statement and sets limits to what the student says.
	9. STUDENT TALK--INITIATION: talk by students which they initiate. Unpredictable statements in response to teacher. Shift from 8 to 9 as student introduces own ideas.
	10. SILENCE OR CONFUSION: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

Figure 6

Categories for Verbal Interaction Analysis

Second

First

	1	2	3	4	5	6	7	8	9	10	Matrix Total
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
Total											
%											

Figure 7

Sample Interaction Matrix

APPENDIX D: Composite Verbal Interaction Data Matrices of the Curricula

TABLE 10

Composite Matrix of Eight Traditional Physics Classes

Category	1	2	3	4	5	6	7	8	9	102
1										
2		18	6	15	56	2		1	11	7
3		1	214	14	63	3	1	2	83	12
4		2	2	101	23	3		110	56	32
5		1	3	105	2,648	24	1	4	117	96
6			2	6	13	46		2	9	14
7				2	5		3		3	3
8		38	20	22	30	3	1	45	7	5
9		53	135	29	54	8	5		407	94
10		3	11	35	107	3	5	7	92	163
Total[a]		116	393	329	2,999	92	16	171	785	426
Percent[b]		2.177	7.377	6.176	56.298	1.727	.300	3.210	14.730	7.996

[a]Total observations, N = 5,327

[b]Expressed in percent of total observations

TABLE 11

Composite Matrix of Ten PSSC Physics Classes

Category	1	2	3	4	5	6	7	8	9	10
1	1							1	1	
2		9	9	10	23			5	7	14
3		2	150	83	57	3	4	10	35	4
4		1	8	187	21	2	3	185	42	110
5	1	6	2	94	1,825	8	1	4	109	152
6				8	4	17		5	11	11
7			2	5	8	1	10	1	12	3
8		25	99	86	28	1	3	251	13	37
9	1	24	67	27	72	6	12	3	252	280
10		10	11	59	164	18	8	79	262	481
Total[a]	3	77	348	559	2,202	56	41	543	744	1,092
Percent[b]	.052	1.359	6.142	10.987	38.870	.988	.723	10.959	13.133	11.928

[a] Total observations, N = 5,665

[b] Expressed in percent of total observations

TABLE 12

Composite Matrix of Eight HPP Classes

Category	1	2	3	4	5	6	7	8	9	10
1	1				1		1			1
2		1	13	69	48	2		3	14	8
3		2	56	11	41	3		6	6	12
4			2	48	12			180	16	17
5	2	3	2	78	3,718	20	7	5	78	96
6				7	7	36	1	1	8	4
7					8		10			2
8		110	33	35	20			40	1	4
9		41	30	5	46	1	1		251	60
10	1	3	1	22	108	2		8	61	207
Total[a]	4	160	137	275	4,009	64	20	243	435	411
Percent[b]	.069	2.778	2.379	4.775	69.624	1.111	.347	4.220	7.554	7.137

[a]Total observations, N = 5,758

[b]Expressed in percent of total observations

APPENDIX E: Composite Verbal Interaction Data Matrices of High and Low Achieving Classes

TABLE 13

Composite Matrix of the Ten Highest Scoring Classes

Category	1	2	3	4	5	6	7	8	9	10
1	2				1		1		1	
2			3	20	28			3	7	10
3		4	143	22	53	3	2	8	59	4
4		1	3	92	12	1		104	22	38
5	1	3	2	71	2,276	11	1	3	67	52
6				5	2	22	1	2	2	5
7				1	7		5	1	5	4
8		38	33	41	14		1	53	12	12
9	1	24	99	10	28	1	9	1	258	46
10	1	1	15	11	66	1	3	29	44	113
Total[a]	5	71	298	273	2,487	39	23	204	477	283
Percent[b]	0.12	1.71	7.16	6.56	59.77	0.94	0.55	4.90	11.46	6.83

[a]Total observations, N = 4,161

[b]Expressed in percent of total observations

TABLE 14

Composite Matrix of the Ten Lowest Scoring Classes

Category	1	2	3	4	5	6	7	8	9	10
1										
2		11	4	7	16	1		3	4	7
3		1	61	8	14		2		25	11
4			4	20	7		2	40	15	24
5		3	4	31	1,571	8	4		61	74
6				2	2	12		3	5	2
7			2	2	7		10		9	2
8		12	10	10	8		1	44	3	14
9		19	34	15	39	3	5	1	119	125
10		7	3	17	92	2	8	11	119	375
Total[a]	0	53	122	112	1,756	26	32	102	360	634
Percent[b]	0.00	1.66	3.82	3.50	54.93	0.81	1.00	3.19	11.26	19.83

[a]Total observations, N = 3,197

[b]Expressed in percent of total observations

APPENDIX F: Correlation Coefficients for Formative Experiences

Table 15

Correlation Coefficients of Formative Experiences, Achievement Measures, and Curricula

	Formative Experiences		
	Academic Aptitude	Sex	Grade Level
TP	-.48	.15	-.02
PSSC	-.33	-.10	.03
HPP	-.42	-.19	-.02
DAPT Pretest	-.16	-.19	.17
DAPT Posttest	-.03	.19	-.19
WISP Pretest	-.09	.13	-.19
WISP Posttest	-.11	.07	-.01
WGCTA Pretest	-.09	.05	.03
WGCTA Posttest	-.09	.06	.03

Note: Pearson's product moment correlation coefficients.

APPENDIX G: Teacher ID Ratio, Pretest and Posttest Mean Scores for Three Instruments by TP, HPP, and PSSC Classroom Groups

TABLE 16

Teacher ID Ratio, Pretest and Posttest Mean Scores for Three Instruments by TP, HPP, and PSSC Classroom Groups

Group Identification Curriculum Teacher No.	Teacher ID Ratio	DAPT Pre	DAPT Post	WISP Pre	WISP Post	WGCTA Pre	WGCTA Post
TP-01	.23	11.74	22.13	58.41	65.87	73.12	76.30
TP-02	.47	13.82	25.03	56.40	61.22	69.42	72.01
TP-02	.47	12.81	24.93	63.00	64.00	70.43	71.93
TP-03	.26	13.51	10.84	53.13	49.84	51.00	39.69
TP-04	.16	19.65	21.75	64.51	66.56	69.31	73.31
TP-05	.23	19.40	27.31	65.81	67.38	65.71	67.08
TP-06	.36	23.11	28.88	62.91	67.23	76.63	78.33
TP-06	.36	17.86	29.07	63.44	66.91	70.39	78.00
TP-07	.15	18.61	26.11	59.43	65.43	70.91	73.49
TP-07	.15	19.65	25.78	63.40	64.92	70.01	72.77
TP-08	.01	18.11	24.22	59.00	67.06	68.00	72.00
HPP-09	.22	16.42	22.31	62.71	62.00	69.72	68.93
HPP-10	.81	12.49	18.03	62.71	63.01	68.91	70.50
HPP-11	.13	17.18	27.06	63.11	66.90	68.32	75.71
HPP-12	.06	16.73	27.52	62.81	64.77	70.81	72.81
HPP-12	.06	22.11	28.03	63.21	65.07	71.18	73.01
HPP-12	.06	17.45	27.99	63.34	64.94	70.01	72.97
HPP-13	.02	19.91	29.00	63.42	68.15	73.81	73.21
HPP-14	.05	17.77	21.92	62.98	60.41	69.83	66.72
HPP-15	.03	12.35	21.79	62.00	63.45	69.35	71.00
HPP-15	.03	14.81	22.00	61.05	62.91	69.35	70.96
HPP-16	.14	15.52	26.22	62.50	66.78	74.52	78.39

TABLE 16 (continued)

Group Identification Curriculum Teacher No.	Teacher ID Ratio	DAPT		WISP		WGCTA	
		Pre	Post	Pre	Post	Pre	Post
PSSC-17	.12	17.32	28.69	64.12	66.32	70.21	72.78
PSSC-18	.11	16.21	27.33	63.44	67.63	77.32	80.38
PSSC-19	.65	15.61	27.67	61.50	64.03	69.00	69.99
PSSC-20	.63	16.31	25.00	67.50	69.84	71.00	77.69
PSSC-21	.03	19.35	27.76	63.51	67.33	73.50	78.02
PSSC-21	.03	18.61	28.41	70.00	66.67	69.81	77.92
PSSC-22	.50	20.91	29.00	63.50	68.11	68.21	75.87
PSSC-22	.50	22.58	29.07	66.00	68.71	71.42	76.40
PSSC-23	.35	17.91	29.67	62.68	66.76	70.34	74.86
PSSC-23	.35	18.09	30.16	63.27	67.03	79.40	75.11
PSSC-24	.27	19.13	23.03	62.61	63.17	70.41	69.33
PSSC-24	.27	22.00	22.99	62.71	62.73	70.00	69.91
PSSC-25	.34	18.79	23.00	63.62	62.70	70.05	70.36
PSSC-25	.34	13.91	22.71	61.30	63.12	64.80	69.88
PSSC-26	.08	15.22	20.70	61.60	64.11	66.12	63.57
PSSC-27	.15	16.11	20.71	62.81	64.10	66.00	63.56

VITA

Vita

Name: T. C. Smith, Jr.

Born: February 18, 1937, Normangee, Texas

Education: B.S., 1959, Sam Houston State College, Huntsville, Texas
M.Ed., 1961, Sam Houston State College, Huntsville, Texas
Ed.D., 1971, University of Houston, Houston, Texas

Additional graduate work:
University of Arizona
Tucson, Arizona
1966-1967

Trinity University
San Antonio, Texas
Summer, 1967

Texas A & M University
College Station, Texas
1964-1966

Professional Experiences: Sam Houston State College
Huntsville, Texas
Instructor, physical science, physics, and mathematics

Blinn Junior College
Brenham, Texas
Chairman, Physics Department

University of Houston
Houston, Texas
Teaching Fellow, Curriculum & Instruction
Instructor, Science Education